AD-A035 222 CALIFORNIA UNIV LOS ANGELES SCHOOL OF ENGINEERING A--ETC F/G 17/2
SOURCE CODING AND ADAPTIVE DATA COMPRESSION FOR COMMUNICATION N--ETC(U)
DEC 76 H S NUSSBAUM N00014-75-C-0609

UNCLASSIFIED UCLA-ENG-76121 NL

1 OF 3
AD
A035222

Research Sponsored by
Office of Naval Research
Under Contract N00014-75-C-0609
and the
National Science Foundation
Under Grant ENG 75-03224



UCLA-ENG-76121
DECEMBER 1976

# SOURCE CODING AND ADAPTIVE DATA COMPRESSION FOR COMMUNICATION NETWORKS

H.S. NUSSBAUM

Principal Investigator: I. RUBIN

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6
SOURCE CODING AND ADAPTIVE DATA COMPRESSION FOR COMMUNICATION NETWORKS

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER
76121

7. AUTHOR(s)
10
Howard S. Nussbaum

8. CONTRACT OR GRANT NUMBER(s)
15
N00014-75-C-0609, NSF-ENG-75-03224

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of California
Los Angeles, California 90024

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR042-291

11. CONTROLLING OFFICE NAME AND ADDRESS
Statistics and Probability Program
Office of Naval Research
Arlington, Virginia 22217

12. REPORT DATE
11
December 1976

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
14
UCLA-ENG-76121

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Reproduction in whole or in part is permitted for any purpose of the United States government.

12 220p.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

| | |
|---|---|
| Source Encoding | Computer Communication Networks |
| Data Compression | Markov Decision Processes |
| Communication Networks | Queueing Theory |
| Rate Distortion Theory | Store-and-Forward Networks |

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Coding under a fidelity criteria of a class of sources which emit randomly occurring messages is investigated. This class of sources models information carrying processes entering into communication networks. Messages emitted by computer terminals, teletypes, vocoders, and other such devices serve as actual examples. For this class of sources the rate distortion function is derived, and source coding and converse source coding theorems are proven. Employing these theorems, an operational definition of the rate

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 637 LB

distortion function in terms of message queueing delay, and transmission delay is presented. This definition relates the rate distortion function with the message network delay which is an important measure of performance of a communication network.

Also, for such a class of sources, which emit randomly occurring messages, an adaptive data compression scheme is investigated. This scheme utilizes observations of the network congestion to determine the amount of compression a message receives, with the object of minimizing the message delay for a given distortion level. Using results from Markov decision theory for the case of a nondenumerable state space and an unbounded cost function, the delay distortion relationship is examined for a tandem channel network which utilizes this adaptive data compression scheme. This relationship represents the trade-off between message delay and fidelity of the reproduction of the source. Using a queueing analysis of the communication system under consideration, numerical results for the delay distortion relationship are provided. Applications of adaptive data compression to analog signals, such as voice or telemetry waveforms transmitted through a radio or satellite channel, are cited as examples.

UCLA-ENG-76121
DECEMBER 1976

# SOURCE CODING AND ADAPTIVE DATA COMPRESSION FOR COMMUNICATION NETWORKS

by

Howard S. Nussbaum

Principal Investigator: Izhak Rubin

Research Sponsored by
Office of Naval Research
Under Contract N00014-75-C-0609
and the
National Science Foundation
Under Grant ENG 75-03224



Department of System Science
School of Engineering and Applied Science
University of California
Los Angeles, California

## ACKNOWLEDGMENTS

It is a great pleasure for the author to acknowledge the guidance provided by Professor Izhak Rubin during the research of this dissertation. Professor Rubin's aid and encouragement throughout the course of this investigation are deeply appreciated. Sincere thanks are extended to Professors Leonard Kleinrock, Thomas M. Liggett, Steven A. Lippman and Kung Yao for serving on the dissertation committee.

Grateful acknowledgment is made to the Hughes Aircraft Company, which provided generous support through a Hughes Staff Doctoral Fellowship. Appreciation is expressed for the spport provided by the National Science Foundation under Grant ENG 75-0324, and by the Office of Naval Research under contract N00014-75-C-0609.

Special appreciation is made to Laraine Simpson for her excellent typing of this manuscript.

## ABSTRACT

Coding under a fidelity criteria of a class of sources which emit randomly occurring messages is investigated. This class of sources models information carrying processes entering into communication networks. Messages emitted by computer terminals, teletypes, vocoders, and other such devices serve as actual examples. For this class of sources the rate distortion function is derived, and source coding and converse souce coding theorems are proven. Employing these theorems, an operational definition of the rate distortion function in terms of message queueing delay, and transmission delay is presented. This definition relates the rate distortion function with the message network delay which is an important measure of performance of a communication network.

Also, for such a class of sources, which emit randomly occurring messages, an adaptive data compression scheme is investigated. This scheme utilizes observations of the network congestion to determine the amount of compression a message receives, with the object of minimizing the message delay for a given distortion level. Using results from Markov decision theory for the case of a nondenumerable state space and an unbounded cost function, the delay distortion relationship is examined for a tandem channel network which utilizes this adaptive data compression scheme. This relationship represents the trade-off between message delay and fidelity of the reproduction of the source. Using a queueing analysis of the communication system under consideration, numerical results for the delay distortion relationship are provided. Applications of




PRECEDING PAGE BLANK-NOT FILMED

adaptive data compression to analog signals, such as voice or telemetry waveforms transmitted through a radio or satellite channel, are cited as examples.

TABLE OF CONTENTS

CONTENTS (continued)

## LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

Results in information theory which pertain to the coding of sources under a fidelity criteria are normally applied to sources which emit signal functions on a regular temporal basis. When considering an information-theoretic approach to the coding of sources found in communication network applications, the existing source models are inadequate. In this dissertation, a source model is considered which more accurately models the information processes conveyed by communication networks. For this source, the corresponding rate distortion function is shown to have an operational meaning in terms of message queueing delays and transmission delays, and an adaptive data compression scheme is developed and optimized.

## 1.1 Problem Statement

A model of a communication system for the conveyance of information from a set of sources to a set of users is depicted in Figure 1.1. This block diagram represents the situation found in communication network applications such as computer and satellite communication networks. Information is generated at the sources by a certain random mechanism. At the source encoder, the messages are processed by a data compressor which generates approximations to the messages. The channel encoder adds redundant information to the latter message approximations to insure reliable transmission through

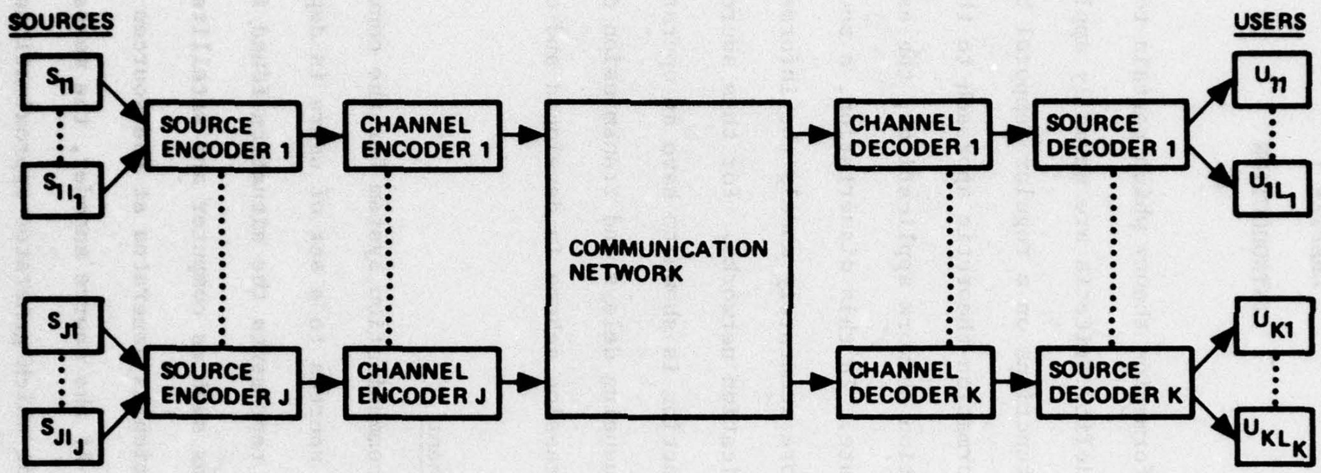



Figure 1.1. Communication System Block Diagram

the communication network. The communication network assumes the responsibility of transmitting the messages from the source to the destination. The network consists of an interconnection of communication channels and buffers. The buffers are used to regulate the bit rate through the channels to insure acceptable message queueing delays. Finally, the channel decoder and source decoder decipher the information transmitted through the network and provide a reproduction of the original message to the users.

The sources are assumed to generate the messages at random instants in time. Examples of such sources are timesharing terminals, stock quotation terminals, vocoders, and others. This randomness in timing is the essential element that differentiates this source from the normal information theoretical model of sources. Thus, the source encoder, as well as other elements of the communication system, must be able to accommodate the resulting temporal randomness. This gives rise to associated queueing delays in the communication system, which along with transmission delays account for the source to user message delays.

The intent of the described communication system is to deliver the messages from the sources to the users as rapidly and reliably as possible. One trade-off, between rapid delivery of the messages and their accurate reproduction, originates in the source encoder and the data compression scheme. As the amount of compressed information in a message is increased, the resulting message delay is reduced. This trade-off between delay and distortion is characterized by the delay distortion relationship. This relationship describes the minimum

average message delay as a function of the average distortion level attainable for a given system configuration and a given class of data compression schemes.

A method of improving the delay distortion relationship is to permit the class of data compression schemes considered in the calculation of this relationship to include schemes which have an element of adaptivity. This allows for alteration of parameters of the data compression scheme as the congestion of the communication network changes, and thereby reducing message delay.

A parameter of interest associated with the delay distortion relationship is the minimum distortion level which achieves a prescribed finite value of the average message delay. This quantity represents the optimal theoretically attainable accuracy of reproduction of the source. It is described by the rate distortion function, an information-theoretic concept. This function yields the idealized relationship between the average rate of tranmission of information in the network, and the associated average distortion.

In this dissertation, a model of sources found in communication networks is presented. Using this source model, the rate distortion function is evaluated, and given an operational definition in terms of message delay. This interpretation is a natural extension of the usual information-theoretic operational definition. Incorporating this source model, a communication system with an adaptive data compression scheme is analyzed. Using Markov decision theory, the structure of the adaptive compression scheme which yields the optimal delay distortion relationship is determined. For network

configurations consisting of a single channel and channels in tandem, the delay distortion relationship is computed through a queueing analysis yielding the average delay of a message in the communication system.

## 1.2 Historical Background

The problem of encoding sources under a fidelity criteria forms a well studied area (rate distortion theory) of information theory. The evaluation of the performance of the encoding schemes in terms of the rate distortion function was introduced by Shannon [1], [2]. This function represents the optimal theoretically attainable rate for a given distortion level. Results incorporating the rate distortion function, see Gallager [3], and Berger [4], are normally applied to source models which have temporal regularity. A typical example of such a source is the discrete time source. This source is assumed to produce each unit of time a letter which is a realization of the underlying stochastic process. As defined, the rate distortion function is independent of delays accrued due to encoding and transmission. An element of coding delay was added to the rate distortion function by Krich and Berger [5]. This was done by introducing a fidelity criteria which depends upon both coding time delays, and the accuracy of the reproduction.

Considering the transmission of information, communication networks designed for the transmission of information in a digital format are of great current interest. The major motivation for this interest is the application of such networks to computer communications

(for example, see the recent books and collections of papers by Davies and Barber [6], Abramson and Kuo [7], Green and Lucky [8], and Chu [9]). In these applications, a network is established to allow subscribers to share resources efficiently. A major operational requirement for such networks is that messages from subscribers be transmitted as reliably and as rapidly as possible. The message processes from the various subscribers enter the network at any specific node. At a node, the processes can be modelled as a single message process generated by a source which emits messages at random instants of time (see, for example Kleinrock [10]). The analysis of the resulting message queueing and transmission delays for such networks can be found, for example, in Kleinrock [10], [11] and Rubin [12]. The analysis of the accuracy of the reproduction of the messages in the network under a fidelity criteria was introduced by Rubin [13]. In that paper, the delay distortion relationship was introduced to display the relationship between the delay of a message in the network, and the amount of distortion accrued by a message when data compressed. The data compression schemes which were considered in [13], only included data compression schemes that do not have any element of adaptivity or decision making capability associated with them.

The analysis of communication networks which contain decision making schemes to control the network can frequently be performed by first modelling the state processes involved by Markov decision processes. Then the network is considered to be a queueing system with a controller and the results of Markov decision theory are used

to analyze the system and optimize it. Results for Markov decision processes with respect to an infinite horizon deal with describing under what conditions a stationary decision policy achieves the minimum discounted cost or the minimum average cost. Studies of such problems can be found, for example, in Blackwell [14], Howard [15], Derman [16], [17], Ross [18], [19], Lippman [20], [21], Reed [22], Harrison [23], and Awate [24]. Only in the papers of Lippman [20], [21], Reed [22], Harrison [23] and Awate [24] are Markov decision processes with unbounded cost functions considered. Furthermore, only Lippman [21] considers such processes with non-denumerable state spaces and unbounded cost functions.

Applications of the results of Markov decision theory to the optimal control of queueing systems are quite extensive. The review articles of Crabill, Gross, and Magaine [25], Sobel [26], and Stidham and Prabhu [27] outline some of the applications to queueing systems. In general, problems dealing with the optimal control of queueing systems consider the state of the system to be the queue size or some other related denumerable state space. Shaw [28] considered a nondenumerable state space, the waiting time, in working with the control of a M/M/1 queueing system with variable service rate. The technique used to solve this latter problem does not, however, rely on the results of Markov decision theory.

## 1.3 Outline of Dissertation

The main results of this dissertation relate to the coding under a fidelity criteria of a class of network sources. This class

of sources are found in communication network applications and the sources do not have an inherent temporal regularity.

In Chapter II, the optimal theoretically attainable performance of a source encoding scheme for such a source is studied. The model of the sources is described in Section 2.1, and the appropriate rate distortion functions for this model are derived in Section 2.2. In Section 2.3, a source coding and a converse source coding theorem are proved, and in Section 2.4, an information transmission theorem and its converse are proved. These theorems establish a natural operational definition of the rate distortion function in terms of source codes which achieve a finite average message delay.

In Chapter III, the relationship between the delay of a message in a network and the distortion accrued from data compression is studied. The delay distortion relationship is introduced in Section 3.1. An adaptive data compression scheme is described in Section 3.2. In Section 3.3, Markov decision processes are introduced and relevant results from Markov decision theory are reviewed. Also, an extension of the results of Markov decision theory to the case of nondenumerable state space and unbounded cost functions is derived in this section. In Section 3.4, the results of Section 3.3 on Markov decision processes are applied to the study of an adaptive data compression scheme for a network consisting of a single channel. The optimal structure of the scheme which achieves the delay distortion relationship is determined.

Considering an adaptive data compression scheme, the delay distortion relationship for a network consisting of a single channel

is further studied in Chapter IV. In Section 4.1, the results of Chapter III are used to specify the delay distortion relationship in terms of an optimization of an appropriate functional. The major properties of this relationship are examined in Section 4.2. In Section 4.3, the functional to be optimized is described in terms of an integral equation. In Section 4.4, a more mathematically tractable method for evaluating the functional, utilizing results from Markov chain theory, is discussed. Numerical results for the delay distortion relationship are presented in Section 4.5.

In Chapter V, a tandem channel communication network is considered and the results of Chapters III and IV on adaptive data compression schemes are extended to this network. The tandem channel network is an accurate model of a satellite communication channel or a radio relay channel. The delay distortion relationship for this network is determined in Section 5.1. The functional form of this relationship is shown to be similar to the functional form of the delay distortion relationship for the single channel network studied in Section 4.1. In Section 5.2, the major properties of this relationship are presented. In Section 5.3, numerical results for the delay distortion relationship of the tandem channel network are presented.

A summary, conclusions, and suggestions for future research are found in Chapter VI.

# CHAPTER II

## SOURCE MODEL AND RATE DISTORTION THEORY FOR COMMUNICATION NETWORKS

The class of sources usually considered in communication network applications generate messages at random instants in time. Examples of such sources are timeshare terminals, stock quotation terminals, and vocoders. In this chapter, a model of these sources is developed, and an information theoretic approach is applied to the coding of these sources. Specifically, the idea of message delay is introduced into the operational definition of the rate distortion function.

In Section 2.1, the model of the sources is described, and measures of fidelity are discussed. The rate distortion function, R(D), for this model is derived and is shown to have a second interpretation in Section 2.2. Source coding, and converse source coding theorems are proved in Section 2.3. These theorems estalish a natural operation definition of R(D) in terms of source codes which insure a finite message delay. Information transmission theorems are established in Section 2.4.

### 2.1 Source Model

Consider the communication system depicted in Figure 1.1. The source model which is considered represents sources that generate messages or letter at random instants of time. This source model is representative of information sources which generate information

in a bursty manner, such as timeshare terminals, stock quotation terminals, and vocoders. Furthermore, many times it is required that scientific instruments make measurements of phenomenons that occur irregularly. Hence, the data collected from the instruments can be modelled as a source that generates messages at random instants in time.

For such sources which generate messages randomly in time, the usual models of sources found in information theory, (see Gallager [3] and Berger [4]) are inadequate to describe the sources. To account for this inadequacy, a model for the superposition of the arrival processes of the messages from the various sources to the source encoder needs to be specified.

The superposition of the arrival processes is assumed to form a renewal point process, where the interarrival times are independent and identically distributed. Let $\{\tau_n, n = 1, 2, 3, \ldots\}$ describe the message interarrival times where $\{\tau_n\}$ is a sequence of non-negative independent random variables with a common distribution function $F_\tau(\cdot)$ and a common expected value, $\lambda^{-1}$, given by

$$\lambda^{-1} = \int_0^\infty \tau d\, F_\tau(\tau) < \infty \quad . \qquad (2.1)$$

Then the sequence of arrival times $\{t_n, n = 0, 1, 2, \ldots\}$ of the renewal process is given by

$$t_n = \sum_{j=0}^{n} \tau_j \quad , \quad n = 0, 1, 2, \ldots \qquad (2.2)$$

Furthermore, define N(T) as the number of messages arriving in the

interval $[0, T]$, $T \geq 0$. Thus, $N(T)$ is given by

$$N(T) = \sup\{n \geq 0: \ t_n \leq T\} \tag{2.3}$$

Messages arriving at different instants of time are drawn from an abstract space, X. Each message, $x \in X$, is considered a letter and the letters are assumed to be independent identically distributed random variables which are statistically independent of the arrival process. The reproduction of $x \in X$ is a letter $y \in Y$ where Y is the abstract reproduction space. Subsequently, an amount of distortion is accrued when $x \in X$ is reproduced by $y \in Y$ which is represented by $\rho(x,y)$. The function $\rho(\cdot,\cdot)$ is a distortion measure and $\rho(x,y)$ is the cost of reproducing x with y.

It is assumed that the users are not concerned with the exact time of arrival of a message. This is reflected by the distortion measure not being a function of the arrival time of the message (occurrence times which need to be transmitted are considered as part of the message). However, users are concerned with the message delay which is an important performance criteria for communication networks. Therefore, the message delay is considered as a separate performance measure, instead of combining the accuracy of signal reproduction and message delay in a single performance criteria as in Kritch and Berger [5].

## 2.2 Rate Distortion Function

The rate distortion function R(D) for sources emitting randomly occurring messages as described in Section 2.1 is now derived. For

these sources, the number of messages arriving in a given time period is random, in contrast to sources which emit messages at regular intervals of time. Using the properties of renewal point processes, it is shown that the rate distortion function for sources emitting randomly occurring messages can be defined in an alternative manner to the definition used for the rate distortion function for sources with temporal regularity.

The computation of R(D) requires that the sample function of the source over the interval [0, T], be represent by

$$\tilde{x}_T = [\{x_n\}, \{t_n\}, N(T)] \tag{2.4}$$

where the sequence of messages is given by

$$\{x_n\} = \{x_n \in X, n = 0, 1, 2, \ldots, N(T)\} \quad , \tag{2.5}$$

the sequence of arrival times is given by

$$\{t_n\} = \{t_n, n = 0, 1, 2, \ldots, N(T)\} \quad , \tag{2.6}$$

and N(T) is the number of messages generated in the interval [0, T]. Similarly, the sample function of the reproduction of the source is represented by

$$\tilde{y}_T = [\{y_n\}, \{\hat{t}_n\}, \hat{N}(T)] \tag{2.7}$$

where the sequence of reproductions of the messages is given by

$$\{y_n\} = \{y_n \in Y, n = 0, 1, 2, \ldots, \hat{N}(T)\} \; , \tag{2.8}$$

the sequence of reproductions of the arrival times is given by

$$\{\hat{t}_n\} = \{\hat{t}_n,\ n = 0, 1, 2, \ldots, \hat{N}(T)\} \tag{2.9}$$

and $\hat{N}(T)$ is the number of messages reproduced. Furthermore, let $\tilde{q}_T$ be the conditional probability measure generating the underlying joint probability measure between $\tilde{x}_T$ and $\tilde{y}_T$.

Using this notation, the average mutual information between the source $\tilde{x}_T$, and the reproduction $\tilde{y}_T$ over the interval [0, T], with respect to $\tilde{q}_T$, is denoted by $I_{\tilde{q}_T}(\tilde{x}_T;\ \tilde{y}_T)$. The distortion accrued when the sample function $\tilde{x}_T$ is reproduced by $\tilde{y}_T$ is denoted $\rho_T(\tilde{x}_T,\ \tilde{y}_T)$. Thus, from Berger [4], the rate distortion function R(D) is defined by

$$R(D) = \lim_{T\to\infty}\ \inf_{\tilde{q}_T \in Q_T(D)} T^{-1}\, I_{\tilde{q}_T}(\tilde{x}_T;\ \tilde{y}_T) \tag{2.10}$$

where

$$Q_T(D) = \{\tilde{q}_T:\ E_{\tilde{q}_T}[\rho_T(\tilde{x}_T,\ \tilde{y}_T)] \le D\} \quad . \tag{2.11}$$

The distortion measure $\rho_T(\cdot,\cdot)$ which is employed, is the time average of the per-message distortion measure $\rho(\cdot,\cdot)$, given in Section 2.1. So, averaging $\rho(x_n,\ y_n)$ over time, the distortion measure to be employed, which is denoted $\rho_T^t(\tilde{x}_T,\ \tilde{y}_T)$, is given by

$$\rho_T^t(\tilde{x}_T,\tilde{y}_T) = \begin{cases} T^{-1}\displaystyle\sum_{n=1}^{N(T)} \rho(x_n,y_n) & ,\ \text{if } \hat{N}(T) \ge N(T) \\ T^{-1}\Big[\displaystyle\sum_{n=0}^{\hat{N}(T)} \rho(x_n,y_n) + (N(T) - \hat{N}(T))\rho_{max}\Big], & \text{if } \hat{N}(T) < N(T) \end{cases} \tag{2.12}$$

where

$$\rho_{max} = \inf_{y \varepsilon Y} E(\rho(x,y)|y) + \varepsilon \tag{2.13}$$

and $\varepsilon > 0$. The quantity $\rho_{max}$ is assumed to be finite, which in general is the case for most distortion measures. The term $(N(T) - \hat{N}(T))\rho_{max}$ in (2.12) is included to insure that the measure is honest in the sense that providing no information about a group of messages is suitably penalized. This honesty is required in the following proposition which establishes an alternative expression for the rate distortion function $R_t(D)$. The latter is defined by incorporating the time average per-message fidelity criteria (2.12) in (2.10) - (2.11) as follows

$$R_t(D) = \lim_{T\to\infty} \inf_{\tilde{q}_T \varepsilon \tilde{Q}_T^t(D)} T^{-1} I_{\tilde{q}_T}(\tilde{x}_T; \tilde{y}_T) \tag{2.14}$$

where

$$Q_T(D) = \{\tilde{q}_T: \; E_{\tilde{q}_T}[\rho_T^t(\tilde{x}_T, \tilde{y}_T)] \le D\} \tag{2.15}$$

Before stating the proposition, the average conditional mutual information must be defined. Let the notation

$$I_{\tilde{q}_T}([\{x_n\}, \{t_n\}]; [\{y_n\}, \{\hat{t}_n\}]|N(T))$$

represent the average conditional mutual information between $[\{x_n\},\{t_n\}]$ and $[\{y_n\},\{\hat{t}_n\}]$ conditioned on $N(T)$ with respect to the marginals generated by $\tilde{q}_T$. Other forms of the average conditional mutual information which are used later have similar means.

**Proposition 2.1**

The rate distortion function $R_t(D)$ defined by (2.14) is given by

$$R_t(D) = \lim_{T\to\infty} \inf_{\tilde{q}_T \in Q_T'(D)} T^{-1} I_{\tilde{q}_T}(\{x_n\}; \{y_n\}|N(T)) \qquad (2.16)$$

where

$$Q_T'(D) = \{\tilde{q}_T: E_{\tilde{q}_T}[\rho_T^t(\tilde{x}_T, \tilde{y}_T)] \leq D,\ E_{\tilde{q}_T}[I(N(T) = \hat{N}(T))] = 1\} \qquad (2.17)$$

and

$$I(N(T) = \hat{N}(T)) = \begin{cases} 1, & \text{if } N(T) = \hat{N}(T) \\ 0, & \text{if } N(T) \neq \hat{N}(T) \end{cases} \qquad (2.18)$$

is the indicator function.

**Proof**

Making repeated use of an equality from information theory, the mutual information function can be expressed as the sum of various average conditional mutual information quantities as follows,

$$\begin{aligned} I_{\tilde{q}_T}(\tilde{x}_T; \tilde{y}_T) = {} & I_{\tilde{q}_T}(N(T); \hat{N}(T)) + I_{\tilde{q}_T}(N(T); [\{y_n\}, \{\hat{t}_n\}]|\hat{N}(T)) \\ & + I_{\tilde{q}_T}([\{x_n\}, \{t_n\}]; \tilde{y}_T|N(T)) \end{aligned} \qquad (2.19)$$

Since mutual information measures are positive,

$$I_{\tilde{q}_T}(\tilde{x}_T; \tilde{y}_T) \geq I_{\tilde{q}_T}([\{x_n\}, \{t_n\}]; \tilde{y}_T|N(T)) \qquad (2.20)$$

Now let $\tilde{y}'_T$ be given by

$$\tilde{y}'_T = [\{y'_n,\ n = 0, 1, 2, \ldots, N(T)\},\ \{\hat{t}'_n,\ n = 0, 1, 2, \ldots, N(T)\},\ N(T)] \tag{2.21}$$

where for $n = 0, 1, 2, \ldots, N(T)$

$$y'_n = \begin{cases} y_c & \text{if} \quad n > \hat{N}(T) \\ y_n & \text{otherwise} \end{cases}, \tag{2.22}$$

and

$$\hat{t}'_n = \begin{cases} t_c & \text{if} \quad n > \hat{N}(T) \\ \hat{t}_n & \text{otherwise} \end{cases}. \tag{2.23}$$

The quantity $y_c \in Y$ is selected such that $E(\rho(x,y_c)|x) < \rho_{max}$ and $t_c$ is an arbitrary constant. Since $y'_T$ contains less information about the original sequences $\{x_n\}$, and $\{t_n\}$ than does the original reproduction $\tilde{y}_T$, then it is clear that

$$I_{\tilde{q}_T}([\{x_n\}, \{t_n\}]; \tilde{y}_T|N(T))$$
$$\geq I_{\tilde{q}_T}([\{x_n\}, \{t_n\}]; [\{y'_n\}, \{\hat{t}'_n\}]|N(T)) \tag{2.24}$$

Furthermore, from (2.12), (2.22) and the definition of $y_c$, it is clear that

$$E_{\tilde{q}_T}(\rho_T(\tilde{x}_T,\tilde{y}_T)) \geq E_{\tilde{q}_T}(\rho_T(\tilde{x}_T,\tilde{y}'_T)) \quad . \tag{2.25}$$

Suppose $\tilde{q}_T \in Q_T(D)$ is generated by $\tilde{x}_T$ and $\tilde{y}_T$, then from (2.21) $\tilde{y}'_T$ is $\tilde{q}_T$ measurable and from (2.25), $E_{\tilde{q}_T}(\tilde{x}_T,\tilde{y}'_T) \leq D$. So, $\tilde{x}_T$ and $\tilde{y}'_T$ generate a conditional probability measure $\tilde{q}'_T \in Q'_T(D)$. Hence letting $\tilde{q}_T$ and

$\tilde{q}'_T$ be as above,

$$I_{\tilde{q}_T}([\{x_n\}, \{t_n\}]; [\{y'_n\}, \{\hat{t}'_n\}] \,|\, N(T))$$

$$= I_{\tilde{q}'_T}([\{x_n\}, \{t_n\}]; [\{y_n\}; \{\hat{t}'_n\}] \,|\, N(T))$$

$$= I_{\tilde{q}'_T}([\{x_n\}, \{t_n\}]; [\{y_n\}, \{\hat{t}_n\}] \,|\, N(T)) \qquad (2.26)$$

where the last equality is a result of $\{y_n\} = \{y'_n\}$ w.p.1 and $\{t_n\} = \{t'_n\}$ w.p.1 under the conditional probability measure $\tilde{q}'_T$. Thus, for every $\tilde{q}_T \in \tilde{Q}^t_T(D)$ there exists a $\tilde{q}'_T \in Q'_T(D)$ such that (2.26) is valid. So, taking the infinimum of (2.26) results in

$$\inf_{\tilde{q}_T \in \tilde{Q}^t_T(D)} I_{\tilde{q}_T}([\{x_n\}, \{t_n\}]; [\{y_n\}, \{\hat{t}'_n\}] \,|\, N(T))$$

$$= \inf_{\tilde{q}'_T \in Q'_T(D)} I_{\tilde{q}'_T}([\{x_n\}, \{t_n\}]; [\{y_n\}, \{\hat{t}_n\}] \,|\, N(T)). \qquad (2.27)$$

Subsequently taking the infimum of (2.20) and using (2.24) and (2.28), results in

$$R_t(D) \geq \lim_{T\to\infty} \inf_{\tilde{q}_T \in Q'_T(D)} T^{-1} I_{\tilde{q}_T}([\{x_n\},\{t_n\}];[\{y_n\},\{\hat{t}_n\}] \,|\, N(T)) \qquad (2.28)$$

An upper bound to $R_t(D)$ is given by

$$R_t(D) \leq \lim_{T\to\infty} \inf_{\tilde{q}_T \in Q'_T(D)} T^{-1} I_{\tilde{q}_T}(\tilde{x}_T;\tilde{y}_T) \quad , \qquad (2.29)$$

since $Q'_T(D)$ is a subset of $\tilde{Q}_T(D)$. From (2.19), if $\tilde{q}_T \in Q'_T(D)$ then

$$I_{\tilde{q}_T}(\tilde{x}_T;\tilde{y}_T) = I_{\tilde{q}_T}(N(T);\hat{N}(T)) + I_{\tilde{q}_T}([\{x_n\},\{t_n\}];[\{y_n\},\{\hat{t}_n\}] \,|\, N(T)) \qquad (2.30)$$

An upper bound to $I_{\tilde{q}_T}(N(T); \tilde{N}(T))$, as shown by Rubin [7], is the entropy of a geometric distribution of mean $E[N(T)]$. Therefore,

$$I_{\tilde{q}_T}(N(T);\tilde{N}(T)) \leq \ell n\ T + \ell n\left(\frac{1 + E[N(T)]}{T}\right) + E[N(T)]\ \ell n(1 + E^{-1}[N(T)]) \quad . \tag{2.31}$$

Since for a renewal point processes

$$\lim_{T\to\infty} E\{\frac{N(T)}{T}\} = \lambda \quad , \tag{2.32}$$

as given by the Elementary Renewal Theorem (Smith [30]), then

$$\lim_{T\to\infty} T^{-1}\ I_{\tilde{q}_T}(N(T);\ \tilde{N}(T)) = 0 \quad . \tag{2.33}$$

Thus, using (2.30) and (2.33) in (2.29) and taking the limit,

$$R_t(D) \leq \lim_{T\to\infty}\ \inf_{\tilde{q}_T \varepsilon Q_T'(D)}\ T^{-1}\ I_{\tilde{q}_T}([\{x_n\},\{t_n\}];[\{y_n\},\{\hat{t}_n\}]|N(T)) \quad . \tag{2.34}$$

Hence, from (2.28), $R_t(D)$ is upper and lower bounded by the same quantity which results in the equality

$$R_t(D) = \lim_{T\to\infty}\ \inf_{\tilde{q}_T \varepsilon Q_T'(D)}\ T^{-1}\ I_{\tilde{q}_T}([\{x_n\},\{t_n\}];[\{y_n\},\{\hat{t}_n\}]|N(T)) \quad . \tag{2.35}$$

To get the final result, repeated use of an equality from information theory is made to get for $\tilde{q}_T \ \varepsilon\ Q_T'(D)$

$$\begin{aligned} I_{\tilde{q}_T}([\{x_n\},\{t_n\}];[\{y_n\}, \{\hat{t}_n\}]|N(T)) &= I_{\tilde{q}_T}(\{x_n\};\{y_n\}|N(T) \\ &+ I_{\tilde{q}_T}(\{x_n\};\{\hat{t}_n\}|N(T), \{y_n\}) + I_{\tilde{q}_T}(\{t_n\};\{y_n\}|N(T), \{x_n\}) \\ &+ I_{\tilde{q}_T}(\{t_n\};\{\hat{t}_n\}|N(T), \{x_n\},\{y_n\}) \quad . \end{aligned} \tag{2.36}$$

Since the sequences $\{t_n\}$ and $\{\hat{t}_n\}$ do not influence the distortion measure, for every $\tilde{q}_T\ \varepsilon\ Q_T'(D)$ there clearly exists a $\tilde{q}_T'\ \varepsilon\ Q_T'(D)$ such

that $\{t_n\}$ and $\{\hat{t}_n\}$ are statistically independent and statistically independent of $\{x_n\}$ and $\{y_n\}$ under $\tilde{q}'_T$ and

$$I_{\tilde{q}_T}(\{x_n\};\ \{y_n\}|N(T)) = I_{\tilde{q}'_T}(\{x_n\};\ \{y_n\}|N(T)) \qquad (2.37)$$

Then from the independence and (2.36)

$$\begin{aligned} I_{\tilde{q}_T}([\{x_n\},\{t_n\}];[\{y_n\},\{\hat{t}_n\}|N(T)) &\geq I_{\tilde{q}_T}(\{x_n\};\{y_n\}|N(T)) \\ &= I_{\tilde{q}'_T}(\{x_n\};\{y_n\}|N(T)) \\ &= I_{\tilde{q}'_T}([\{x_n\},\{t_n\}];[\{y_n\},\{\hat{t}_n\}]|N(T)) \end{aligned} \qquad (2.38)$$

Substituting (2.38) into (2.35), taking the infimum over $\tilde{q}_T \in \tilde{Q}'_T(D)$, and noting that for every $\tilde{q}_T \in Q'(D)$ there exists a $\tilde{q}'_T(D) \in Q'_T(D)$ that satisfies (2.38), the final result (2.14) is achieved.

<u>Q.E.D</u>

Let the rate distortion function, $r(D)$, for a single letter be defined by

$$r(D) = \inf_{q \,\varepsilon\, Q(D)} I_q(X;Y) \quad , \quad 0 \leq D \leq D_{max} < \rho_{max} \qquad (2.39)$$

where

$$Q = \{q:\ E_q(\rho(x,y) \leq D\}\ . \qquad (2.40)$$

Then, the rate distortion function of a source emitting randomly occurring messages is obtained in Proposition 2.2 in terms of $r(D)$. First note that from information theory $r(D)$ is convex and that

$r(D)$ is assumed continuous.

**Proposition 2.2**

The rate distortion function $R_t(D)$ of a source emitting randomly occurring messages with intensity $\lambda$ (as described in Section 2.1) under a time averaged per-message distortion measure is given by

$$R_t(D) = \lambda\, r(D/\lambda) \tag{2.41}$$

where $r(D)$ is given by (2.39).

**Proof**

Let $q_T^N$ be the conditional probability measure generating the underlying joint probability measure between $\tilde{x}_T$ and $\tilde{y}_T$ conditioned on the event $N(T) = N$ occurring. Furthermore, let $I_{q_T^N}(\{x_n\};\{y_n\}|N(T))$ denote the average mutual information between $\{x_n\}$ and $\{y_n\}$ given $N(T) = N$ with respect to $q_T^N$.

Now define for all $N \geq 0$

$$Q_T^N(D) = \Big\{q_T^N:\ E_{q_T^N}[\rho_T(\{x_n\},\{y_n\})|N(T) = N] \leq D \quad \text{and} \quad E_{q_T^N}[I(N(T) = \hat{N}(T)|N(T) = N] = 1\Big\} \tag{2.42}$$

and

$$\mathcal{D}_T(D) = \Big\{\{D_N \geq 0,\ N = 0,1,2,\ldots\}:\ E(D_{N(T)}) \leq D\Big\} \tag{2.43}$$

where $E(D_{N(T)})$ is the expected value of the random variable $D_{N(T)}$. For any $\tilde{q}_T \in Q_T'(D)$ it is clear that the sequence

$$D_N = E_{\tilde{q}_T}[\rho_T(\{x_n\},\{y_n\})|N(T) = N] \qquad N = 0, 1, 2, \ldots \tag{2.44}$$

satisfies $\{D_N\} \in \mathcal{D}_T(D)$. Furthermore, $\tilde{q}_T$ can be decomposed in the following manner:

$$\tilde{q}_T = \sum_{N=0}^{\infty} q_T^N \, I(N(T) = N) \tag{2.45}$$

where for all $N \geq 0$, $q_T^N \in Q_T^N(D_N)$ and $D_N$ satisfy (2.44).

In a similar manner, if $\{D_N\} \in \mathcal{D}_T(D)$ and $q_T^N \in Q_T^N(D_N)$ for all $N \geq 0$, then $\tilde{q}_T$ is given by (2.45) and $\tilde{q}_T \in \tilde{Q}_T'(D)$. Hence, it is easily shown that

$$\inf_{\tilde{q}_T \in Q_T'(D)} I_{\tilde{q}_T}(\{x_n\};\{y_n\}|N(T)) = \inf_{\substack{\{D_N\} \in \mathcal{D}_T(D) \\ q_T^N \in Q_T^N(D_N),\ N \geq 0}} E\{I_{q_T^{N(T)}}(\{x_n\};\{y_n\}|N(T))\} \tag{2.46}$$

So, applying Fatou's Lemma (see Rudin [31]) to (2.46), the following bound is obtained,

$$\inf_{\tilde{q}_T \in Q_T'(D)} I_{\tilde{q}_T}(\{x_n\};\{y_n\}|N(T)) \geq \inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{\inf_{q_T^{N(T)} \in Q_T^{N(T)}(D_{N(T)})} I_{q_T^{N(T)}}(\{x_n\};\{y_n\}|N(T))\} \tag{2.47}$$

Now let $\varepsilon$ be an arbitrary positive number. Now for every D and N there exists a $\hat{q}_T^N \in Q_T^N(D)$ such that

$$I_{\hat{q}_T^N}(\{x_n\};\{y_n\}|N(T)) = \inf_{q_T^N \in Q_T^N(D)} I_{q_T^N}(\{x_n\};\{y_n\}|N(T)) + \frac{\varepsilon}{2} \tag{2.48}$$

and a sequence $\{\hat{D}_N\} \in \mathcal{D}_T(D)$ such that

$$E\{\inf_{q_T^{N(T)} \in Q_T^{N(T)}(\hat{D}_{N(T)})} I_{q_T^{N(T)}}(\{x_n\};\{y_n\}|N(T))\}$$

$$= \inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{\inf_{q_T^{N(T)} \in Q_T^{N(T)}(D_{N(T)})} I_{q_T^{N(T)}}(\{x_n\};\{y_n\}|N(T))\} + \varepsilon/2 \tag{2.49}$$

Hence, using $\{\hat{D}_N\}$ that satisfies (2.49) and $\hat{q}_T^N \in Q_T^N(D_N)$ that satisfies (2.48), equation (2.45) determines $\tilde{q}_T \in Q_T'(D)$ that satisfies

$$I_{\tilde{q}_T}(\{x_n\};\{y_n\}|N(T))$$

$$= \inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{\inf_{q_T^{N(T)} \in Q_T^{N(T)}(D_{N(T)})} I_{q_T^{N(T)}}(\{x_n\};\{y_n\}|N(T))\} + \varepsilon \tag{2.50}$$

Since $\varepsilon$ is arbitrary, taking the infimum of (2.50) over $\tilde{q}_T \in Q_T'(D)$, the reverse inequality of (2.47) is produced, and

$$\inf_{\tilde{q}_T \in Q_T'(D)} I_{\tilde{q}_T}(\{x_n\};\{y_n\}|N(T))$$

$$= \inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{\inf_{q_T^{N(T)} \in Q_T^{N(T)}(D_N)} I_{q_T^{N(T)}}(\{x_n\};\{y_n\}|N(T))\} \tag{2.51}$$

Now $\inf_{q_T^N \in Q_T^N(D)} I_{q_T^N}(\{x_n\};\{y_n\}|N(T))$ is equal to the $N^{th}$ extension of the rate distortion function, with respect to a per-letter distortion measure $\rho(\cdot,\cdot)$, of a source with letters $\{x_n \in X, n = 0, 1, 2, \ldots, N\}$ which are independent and identically distributed. So, it is clear that

$$\inf_{q_T^N \in Q_T^N(D)} I_{q_T^N}(\{x_n\};\{y_n\}|N(T)) = N\, r(DT/N) \tag{2.52}$$

Using the result from Proposition 2.1, and substituting (2.51) and (2.52) into (2.16), it is concluded that

$$R_t(D) = \lim_{T\to\infty} \inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{(N(T)/T)\, r(D_{N(T)} T/N(T))\} \tag{2.53}$$

Since r(D) is a convex function of D, Jensen's inequality gives

$$E\{(N(T)/T)\, r(D_{N(T)} T/N(T))\} \geq E\{(N(T)/T)\, r(D/(T^{-1}N(T))\} \tag{2.54}$$

which is valid for all $\{D_N\} \in \mathcal{D}_T(D)$ and all $T > 0$. Hence, from (2.52) the sequence $\{D_N\}$ given by

$$D_N = D/E(T^{-1}N(T)), \quad N = 0, 1, 2, \ldots$$

is an element of $\mathcal{D}_T(D)$, and achieves the infimum in (2.51) for all $T \geq 0$. This implies that

$$R_t(D) = \lim_{T\to\infty} E(N(T)/T)\, r(D/E(T^{-1}N(T)) \tag{2.55}$$

Finally, using (2.31) and noting the continuity of r(D), the required result (2.41) is achieved.

<u>Q.E.D</u>

In evaluating the rate distortion function (2.14) a limit is taken with respect to the time period T. In an analogous manner a rate distortion relationship, $\hat{R}_t(D)$, can be defined by taking a limit with respect to the number of messages. For that purpose,

define, in an analogous manner to $\tilde{x}_T$ and $\tilde{y}_T$, $\tilde{x}_N$ and $\tilde{y}_N$ by

$$\tilde{x}_N = [\{x_n\}_N, \{t_n\}_N] \tag{2.56}$$

and

$$\tilde{y}_N = [\{y_n\}_N, \{\hat{t}_n\}_N] \tag{2.57}$$

where $\{x_n\}_N$, $\{y_n\}_N$, $\{t_n\}_N$ and $\{\hat{t}_n\}_N$ are sequences representing the first N messages, reproductions, times of arrivals of messages and reproductions of the times of arrivals, respectively. Then, let $\rho_N^t(\cdot,\cdot)$ be the distortion measure given by

$$\rho_N^t(\tilde{x}_N, \tilde{y}_N) = \frac{\lambda}{N} \sum_{n=0}^{N-1} \rho(x_n, y_n) \quad , \tag{2.58}$$

which is the analogous to the distortion measure given by (2.12). Now $\hat{R}_t(D)$ is defined, with respect to the distortion measure $\rho_N^t(\cdot,\cdot)$, by

$$\hat{R}_t(D) = \lim_{N\to\infty} \inf_{\tilde{q}_N \in Q_N(D)} (\lambda/N)\, I_{\tilde{q}_N}(\tilde{x}_N; \tilde{y}_N) \tag{2.59}$$

where $\tilde{q}_N$ is a conditional probability measure and

$$Q_N = \{\tilde{q}_N: \ E_{\tilde{q}_N}[(\lambda/N) \sum_{n=1}^{N-1} \rho(x_n, y_n)] \le D\} \ . \tag{2.60}$$

The following proposition asserts that taking the limit with respect to time ($T \to \infty$), according to (2.14), or taking the limit with respect to the number of messages ($N \to \infty$), according to (2.58), the same rate distortion relationship is achieved.

**Proposition 2.3**

The rate distortion relationship $\hat{R}_t(D)$ defined by (2.59) is equivalent to the rate distortion function $R_t(D)$ defined by (2.14) for sources emitting randomly occurring messages described in Section 2.1. Thus, for all $D \geq 0$

$$\hat{R}_t(D) = R_t(D) \quad . \tag{2.61}$$

**Proof**

Using a result similar to (2.38), $\hat{R}_t(D)$ is given by

$$\hat{R}_t(D) = \lim_{N\to\infty} \inf_{\tilde{q}_N \in Q_N(D)} (\lambda/N)\, I_{q_N}(\{x_n\}_N; \{y_n\}_N) \tag{2.62}$$

Now $\inf_{\tilde{q}_N \in Q_N(D)} I_{q_N}(\{x_n\}_N; \{y_n\}_N)$ is given by the $N^{th}$ extension of the rate distortion function of a source with letters $x_n \in X$ which are independent and identically distributed. Hence, in a manner similar to (2.52)

$$\inf_{\tilde{q}_N \in Q_N(D)} I_{q_N}(\{x_n\}_N; \{y_n\}_N) = N\, r(D/\lambda) \tag{2.63}$$

and

$$\begin{aligned} \hat{R}_t(D) &= \lambda\, r(D/\lambda) \\ &= R_t(D) \end{aligned} \tag{2.64}$$

by Proposition 2.2.

Q.E.D

Another distortion measure, that is of a similar form to $\rho_T^t(\tilde{x}_T, \tilde{y}_T)$, is the per-message distortion measure given by

$$\rho_T^m(\tilde{x}_T,\tilde{y}_T) = \begin{cases} N(T)^{-1} \sum_{n=1}^{N(T)} \rho(x_n, y_n) & \text{, if } N(T) < \hat{N}(T) \\ N(T)^{-1}[\sum_{n=1}^{N(T)} \rho(x_n,y_n)+(N(T)-\hat{N}(T))\rho_{max}] & \text{, if } N(T) \geq \hat{N}(T) \end{cases} \tag{2.65}$$

The rate distortion function, $R_m(D)$, under this distortion measure is defined as

$$R_m(D) = \lim_{T\to\infty} \inf_{\tilde{q}_T \varepsilon \tilde{Q}_T^m(D)} I_{\tilde{q}_T}(\tilde{x}_T; \tilde{y}_T) \tag{2.66}$$

where

$$Q_T^m(D) = \{\tilde{q}_T: \ E_{\tilde{q}_T}(\rho_T^m(\tilde{x}_T, \tilde{y}_T)) \leq D\} \ . \tag{2.67}$$

In a similar manner to the definition of $\hat{R}_t(D)$, a rate distortion relationship $\hat{R}_m(D)$ can be defined as

$$\hat{R}_m(D) = \lim_{N\to\infty} \inf_{\tilde{q}_N \ \varepsilon \ Q_N^m(D)} (\lambda/N) \ I_{\tilde{q}_N}(\tilde{x}_N; \tilde{y}_N) \tag{2.68}$$

where

$$Q_N^m(D) = \{\tilde{q}_N: E_{\tilde{q}_N}(N^{-1} \sum_{n=0}^{N-1} \rho(x_n, y_n)) \leq D\} \ . \tag{2.69}$$

Proposition 2.4 shows the equivalence between $R_m(D)$ and $\hat{R}_m(D)$.

Proposition 2.4

For sources emitting randomly occurring messages which are described in Section 2.1, the rate distortion function, $R_m(D)$, defined by (2.66) and the rate distortion relationship, $\hat{R}_m(D)$, defined by (2.68) are given by

$$R_m(D) = \hat{R}_m(D) = \lambda\ r(D) \tag{2.70}$$

where r(D) is given by (2.39).

Proof

First, $\hat{R}_t(D)$ as defined in (2.59) is related to $\hat{R}_m(D)$ by

$$\hat{R}_m(D) = \hat{R}_t(D\lambda) \quad . \tag{2.71}$$

So, from Proposition 2.2 and 2.3

$$\hat{R}_m(D) = \lambda\ r(D) \tag{2.72}$$

Now using the same reasoning which is used to develop (2.53), $R_m(D)$ is expressed as

$$R_m(D) = \lim_{T\to\infty} \inf_{\{D_N\}\varepsilon\mathcal{D}_T(D)} E\{N(T)/T\ r(D_{N(T)})\} \tag{2.73}$$

where $\mathcal{D}_T(D)$ is given by (2.43). To evaluate (2.73) define the set $A_T(\varepsilon)$, $\varepsilon > 0$, by

$$A_T(\varepsilon) = \{N\text{: } N \text{ is a positive integer, } \left|\frac{N}{T} - \lambda\right| < \varepsilon\} \ , \tag{2.74}$$

and the indicator function $I_{A_T(\varepsilon)}(\cdot)$ by

$$I_{A_T(\varepsilon)}(N) = \begin{cases} 1 & \text{if} \quad N \in A_T(\varepsilon) \\ 0 & \text{otherwise} \end{cases} \qquad (2.75)$$

Letting $P_T(\varepsilon)$ be given by

$$P_T(\varepsilon) = \Pr(N(T) \notin A_T(\varepsilon)) \quad , \qquad (2.76)$$

the following upper and lower bounds are generated when $\varepsilon$ is chosen such that $P_T(\varepsilon) < D/D_{max}$:

$$\inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{(N(T)/T)\, r(D_{N(T)}) I_{A_T(\varepsilon)}(N(T))\}$$

$$\leq \inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{(N(T)/T)\, r(D_{N(T)})\}$$

$$\leq \inf_{\{D_N\} \in \mathcal{D}_T(D-D_{max}\, P_T(\varepsilon))} E\{(N(T)/T)\, r(D_{N(T)}) I_{A_T(\varepsilon)}(N(T))\}$$

$$+ r(D_{max}) \Pr(A_T(\varepsilon)) \quad . \qquad (2.77)$$

Since for $N(T) \in A_T(\varepsilon)$, it is easily established that

$$(\lambda - \varepsilon) \leq N(T)/T \leq (\lambda + \varepsilon) \quad . \qquad (2.78)$$

So, substituting (2.78) into (2.77) and using Jensen's inequality, it is easily seen that

$$(\lambda-\varepsilon)\ (1 - P_T(\varepsilon))\ r(D) \leq \inf_{\{D_N\} \in \mathcal{D}_T(D)} E\{(N(T)/T)\ r(D_{N(T)})\}$$

$$\leq (\lambda+\varepsilon)\ (1 - P_T(\varepsilon))\ r(D - D_{max}\, P_T(\varepsilon)) + r(D_{max})\, P_T(\varepsilon) \qquad (2.79)$$

Using a result from renewal theory (see Smith [30])

$$\lim_{T\to\infty} N(T)/T = \lambda \qquad \text{w.p.1}, \tag{2.80}$$

it is easy to show $\lim_{T\to\infty} P_T(\varepsilon) = 0$ for all $\varepsilon > 0$. Hence, the limit with respect to T of (2.79) is given by

$$(\lambda - \varepsilon)\, r(D) \leq R_m(D) \leq (\lambda + \varepsilon)\, r(D) \quad . \tag{2.81}$$

Since $\varepsilon$ is an arbitrary positive number, equation (2.70) is obtained.

Q.E.D

Thus, the limiting procedure used in the definition of the rate distortion function can be applied in terms of time ($\lim_{T\to\infty}$) or number of messages ($\lim_{N\to\infty}$). The coding theorems in Section 2.3 establish the importance of this result. They establish that source coding procedures can be used to encode blocks composed of a fixed number of messages, rather than encode blocks of fixed time durations.

## 2.3 Source Coding Theorems

The source coding theorems established in information theory (see, for example, Gallager [3], Berger [4]) provide an operational definition of the rate distortion function, R(D). For a wide class of sources, such theorems establish the existence of a block code, used for encoding the source, whose rate is $R(D) + \varepsilon$. Furthermore, this code yields a distortion level less than $D + \varepsilon$ where $\varepsilon$ is an arbitrary positive number. Converse coding theorems establish that no block code, used for source encoding, exists that

has a rate less than R(D) and yields a distortion level not larger than D. When considering sources that emit randomly occurring messages, which are described in Section 2.1, the above type of source coding theorems for rate distortion functions $R_t(D)$ defined by (2.14) and $R_m(D)$ defined by (2.66) can be proved. These theorems do not make use of the nature of the communication network involved. In this section, another operational definition of the rate distortion function for sources emitting randomly occurring messages is shown to exist by proving a different set of source coding theorems which use concepts relevent to the operation of communication networks.

First, consider the normal scheme of encoding blocks of duration T of the source described in Section 2.1 with a block code. For the implementation of such a coding scheme define a set $\mathcal{N}(T)$ consisting of values of the number of arrivals, N(T), in a duration T such that

$$P_r(N(T) \notin \mathcal{N}(T)) < \varepsilon \tag{2.82}$$

where $\varepsilon > 0$. For each $N(T) \in \mathcal{N}(T)$ a different block code is used to encode the messages, and also, N(T) must be encoded in a distortion free manner. Then, a source coding theorem can be derived in the usual manner by letting the block length T become arbitrarily large.

The disadvantages of such an encoding scheme are that an extremely large code book is required which adds complexity, and overhead information must be included to encode N(T). Therefore, it would be advantageous to be able to encode in a block a constant number of messages N, and thus, reduce the size of the code book and remove the requirement of encoding N(T). Then the question of how to define

the rate of such a message must be answered.

This question is answered by observing that in any communication network buffers are placed at various positions in the network. The purpose of the buffers is to smooth the flow of data by storing the randomly arriving data streams prior to transmission. Making use of a buffer which follows the source encoding procedure, the encoded version of the source is assumed to be placed in a buffer. The information stored in the buffer is further assumed to be released or transmitted from the buffer at a constant rate, r, when the buffer is not empty. So, two measures of the performance of such a block encoding scheme, which encodes a source using a fixed number of messages, are r and the average distortion level the code yields.

A third element of the performance of such a message block encoding scheme is the average message delay, $\gamma$, of a message due to source encoding. The quantity $\gamma$ represents the average time it takes a message after being encoded to leave the buffer which follows the source encoder relative to the time of arrival of the message. The three components of $\gamma$ are the delay introduced in collecting a fixed block of messages, the queueing delay in the buffer, and the time to transmit the encoded message out of the buffer. It is clear that $\gamma$ must be finite to insure the realizability of the encoding scheme. Since without this restriction, the rate r could be arbitrarily small, and the code would still yield a distortion level D, but information would be accumulating in the buffer causing the limiting average message delay to be unbounded.

Using this concept of delay, Theorem 2.1 and Theorem 2.2 establishes an operational definition of the rate distortion function $R_t(D)$ or $R_m(D)$. This definition states that $R_t(D)$ or $R_m(D)$ is the smallest rate of transmission out of the buffer such that the limiting average message delay, $\gamma$, is finite and the code used yields a distortion level not exceeding D.

Theorem 2.1 (Source Coding Theorem for Network Message Sources)

Consider the network message source described in Section 2.1 with finite first and second moments of the interarrival time. Let $\rho(\cdot,\cdot)$ be a given single letter distortion measure.

Let $R_m(\cdot)$ of equation (2.66) denote the rate distortion function of the source with respect to the average message distortion measure (2.65). Then given any $\varepsilon > 0$ and any $D \geq 0$, there exists a $(D + \varepsilon)$ admissible message block code and a rate of transmission $r \leq R(D) + \varepsilon$ which guarantees a finite limiting average message delay.

Proof

Consider (n, m) message block coding schemes which code n messages using a code containing $2^m$ code words. From the source coding theorem for sources from abstract alphabets in Berger [4], it is shown that for any $\varepsilon > 0$, and any $D > 0$ there exists m and n such that a (n, m) coding scheme is $(D + \varepsilon)$ admissible and

$$m < n[r(D) + \frac{\varepsilon}{\lambda}] \tag{2.83}$$

where r(D) is given by (2.39).

Using a (n, m) coding scheme, the buffer which receives the code words behaves as the queue in a GI/D/1 queueing system whose interarrival times are the corresponding sum of the arrival times of n messages, and whose service times are equal to m/r. Fixing $\varepsilon > 0$, and $D \geq 0$, and letting

$$r = R_m(D) + \varepsilon = \lambda\, r(D) + \varepsilon\ , \qquad (2.84)$$

for some message block coding scheme, (n, m), which is $(D + \varepsilon)$ admissible, the ratio $\rho$ of the mean service time, m/r, to the mean interarrival time, $n\lambda^{-1}$, is bounded using (2.83), and (2.84) by

$$\rho = \frac{\lambda/n}{r/m} < \frac{\lambda[r(D) + \varepsilon/\lambda]}{\lambda r(D) + \varepsilon} < 1\ . \qquad (2.85)$$

Since $\rho < 1$, the result of Kingman [32] for a GI/G/1 queue can be used to bound the limiting average queueing delay, $\bar{W}$, of a message in the buffer by

$$\bar{W} \leq \frac{n\sigma_T^{\,2} + \sigma_S^{\,2}}{2n\lambda^{-1}(1-\rho)}\ . \qquad (2.86)$$

The quantity $\sigma_T^{\,2}$ is the variance of the message interarrival time and the quantity $\sigma_S^{\,2}$ is the variance of the service time which is equal to zero for the case under consideration since the service time is deterministic. This implies that $\bar{W}$ is finite since the first two moments of the message interarrival time are finite. Furthermore, for any (n, m) coding scheme and $r > 0$, the transmission delay and the delay associated with collecting n messages are finite. Consequently, a message block code can be found whose limiting average

delay $\gamma$, which is the sum of the three delay components, is finite.

Q.E.D.

The corresponding converse of the theorem is stated in Theorem 2.2.

Theorem 2.2 (Converse Source Coding Theorem for Network Message Sources)

Under the conditions of Theorem 2.1, there exists no D-admissible message block code with an associated transmission rate, $r \leq R_m(D)$ which guarantees a finite limiting average message delay.

Proof

As in the proof of Theorem 2.1, consider a message block coding scheme which codes n messages using a code containing $2^m$ code words. From the converse to the source coding theorem for sources from abstract alphabets in Berger [4], it is established that there exists no D-admissible (n, m) block code encoding n messages into m bits with

$$m \leq n\, r(D) \tag{2.87}$$

Now assume that there exists a D-admissible (n, m) code with

$$m > n\, r(D) \quad . \tag{2.88}$$

The buffer that follows the encoder behaves as a queue in a GI/D/1 queueing system with mean interarrival time $n\lambda^{-1}$ and with mean service time m/r. Letting

$$r \leq R_m(D) \quad , \tag{2.89}$$

the ratio $\rho$ of the mean service time to the mean interarrival time is bounded using (2.88) and (2.89) by

$$\rho = \frac{m/r}{n/\lambda} > \frac{\lambda\, r(D)}{R_m(D)} \tag{2.90}$$

Consequently, substituting the value of $R_m(D)$ from (2.70), $\rho > 1$ is obtained. From Lindley's theorem, Lindley [33], for $\rho > 1$, the limiting average queueing delay is infinite. Therefore, $\gamma$ is infinite.

Q.E.D

These results extend readily to the time average per-message distortion measure and the associated rate distortion function $R_t(D)$ defined in (2.41). Corollary 2.1 establishes the source coding theorem and its converse for this case.

Corollary 2.1

If in Theorems 2.1 and 2.2 the rate distortion function is replaced with the rate distortion function, $R_t(D)$, defined with respect to the time average per-message distortion measure (2.12), then the theorems remain valid.

Proof

Since Theorems 2.1 and 2.2 are valid for all values of distortion, they must be valid for all distortion values of the form $D/\lambda$ where $D \geq 0$. Now a $D/\lambda$-admissible message block code is a time average

D-admissible message block code. So, by Propositions 2.2 and 2.4

$$R_t(D) = R_m(D/\lambda) \tag{2.91}$$

and the result follows.

<u>Q.E.D</u>

From the proofs of the theorems a number of observations can be made. As the excess distortion approaches zero ($\varepsilon \to 0$), the limit of the average message delay becomes arbitrarily large. Furthermore, as $\varepsilon \to 0$, the percentage of time the buffer is empty becomes arbitrarily small, and idle periods in transmission occur infrequently. Hence, information is being transmitted from the buffer at a uniform rate, $r = R(D) + \varepsilon$ almost all the time.

## 2.4 Information Transmission Theorems

The source coding theorem and its converse stated in Section 2.3 establish bounds on the performance of source encoders. These theorems together with the noisy-channel coding theorem, Gallager [3] and its converse, establish bounds on the performance of a communication system. The communication system to be considered is depicted in Figure 2.1 where the system is restricted to a single source-user pair and the communication network consists of a single discrete memoryless channel with channel capacity, C. The information transmission theorem and its converse theorem which follows, specifies for the communication system and a source emitting randomly occurring messages, the theoretical limit of the fidelity of the source

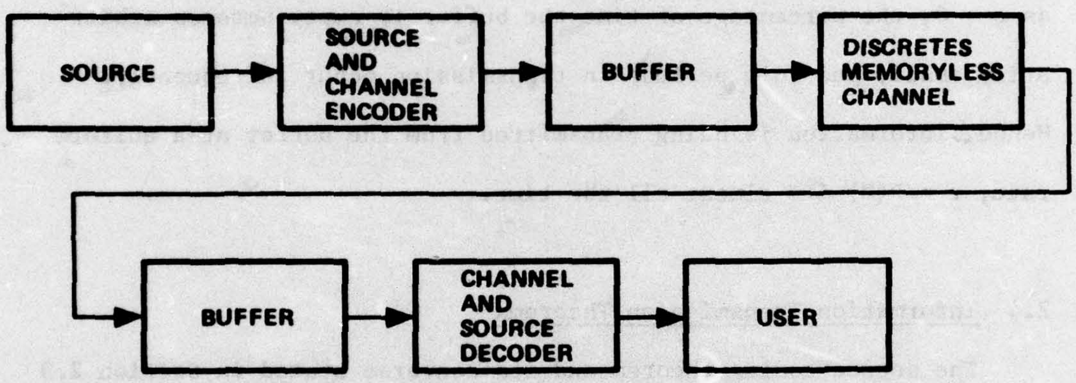



**Figure 2.1. Communication System Employing a Discrete Memoryless Channel**

reproduction at the user.

Theorem 2.3 (Information Transmission Theorem for Network Message Sources)

Let the single lettered distortion measure $\rho(\cdot,\cdot)$ be bounded by P. Then for all $\varepsilon > 0$, for the source of Theorem 2.1 and for any discrete memoryless channel of capacity C, if $C > R_m(D) + \varepsilon$, there exists a message block code which can reproduce the source with fidelity $(D + \varepsilon)$ at the output of the channel and can guarantee a finite limiting average message delay.

Proof

Assume the discrete memoryless channel can transmit one of a finite number of channel symbols and the rate of transmission is r[channel symbols/sec]. Now consider the coding scheme which first uses a message block code to encode the source, and then the resulting code words are channel encoded using a block code and sent to a buffer to await transmission. Certainly, the result of the two coding operations is a message block code.

To evaluate the performance of such a scheme, the source coding theorem and the noisy channel coding theorem are used. Assume $\varepsilon > 0$ and $D > 0$. Using Berger's [4] result for coding of sources with abstract alphabets, there exists a $(D + \varepsilon/2)$-admissible message block code, (n, m), which encodes n messages and has $2^m$ code words where

$$m < n[r(D) + \varepsilon/2\lambda] = \frac{n}{\lambda}\,[R_m(D) + \frac{\varepsilon}{2}]. \qquad (2.92)$$

Since the sequence of arriving messages are independent and identically distributed, the sequence of code words are independent and identically distributed. Thus, the *noisy-channel coding theorem*, Gallager [3], for a discrete memoryless channel applies. It states that a block code, (N, M), can be found that encodes N code words in a distortion free manner into M channel synbols such that

$$C - \frac{\varepsilon}{2} < \frac{m \cdot N}{M} \cdot r \tag{2.93}$$

and

$$P_r(\text{block error at the receiver}) < \varepsilon/2P \tag{2.94}$$

when the channel symbols are transmitted at rate r. Therefore, the scheme utilized is $(D + \varepsilon)$-admissible, since the sum of the distortion due to source encoding and due to channel errors is less than $D + \varepsilon$. Now the channel and buffer represent a GI/D/1 queueing system where a channel code word is the commodity to be serviced. The inter-arrival time of channel code words is the sum of nN message inter-arrival periods. Thus, for this system the service time is $Mr^{-1}$ (a deterministic value), the mean interarrival time is $nN\lambda^{-1}$, and the ratio $\rho$ of these quantities is bounded using (2.93) by

$$\rho = \frac{Mr^{-1}}{nN\lambda^{-1}} < \frac{R_m(D) + \varepsilon/2}{C - \varepsilon/2} \tag{2.95}$$

Since

$$R_m(D) + \varepsilon < C \quad , \tag{2.96}$$

$\rho < 1$ is obtained. As in Theorem 2.1, the condition that the second

moment of the message interarrival is finite implies that the limiting average message delay is finite for a message block code.

Q.E.D

Similarly, the converse theorem states the following.

Theorem 2.4 (Converse Information Transmission Theorem for Network Message Sources)

It is impossible to reproduce the source of Theorem 2.2 with fidelity D at the receiving end of any discrete memoryless channel of capacity $C \leq R_m(D)$ using a message block code which achieves a finite limiting average message delay.

**Proof**

As in Theorem 2.3, assume the discrete memoryless channel can transmit one of a finite number of channel symbols and the rate of transmission is r[channel symbols/sec]. Now consider a block code which codes n messages using a code containing $2^N$ code words where each code word is represented by M channel symbols. From the converse to the source coding theorem for sources from abstract alphabets (Berger [4]), for the block code to be D-admissible, it must satisfy

$$N \geq n\, r(D) = \frac{n}{\lambda} R_m(D) \tag{2.97}$$

Now the buffer and channel form a GI/D/1 queueing system with the transmission time of a code word being equal to the service time of the queueing system. For this system, the service time is $Mr^{-1}$,

the mean interarrival time is $n\lambda^{-1}$, and the ratio $\rho$ of these quantities is given by

$$\rho = \frac{Mr^{-1}}{n\lambda^{-1}} \quad . \tag{2.98}$$

There are two cases to be considered. First, the case $\rho \geq 1$, which implies

$$Mr^{-1} \geq n\lambda^{-1} \quad . \tag{2.99}$$

By Lindley's Theorem (Lindley [33]), the limiting average message queueing delay is infinite, which implies the limiting average message delay is infinite. In the second case, $\rho < 1$ which implies

$$Mr^{-1} < n\lambda^{-1} \quad . \tag{2.100}$$

Now the amount of information transmitted per unit time is, R, given by

$$R = \frac{\frac{n}{\lambda} R_m(D)}{Mr^{-1}} \tag{2.101}$$

and by (2.100)

$$R > R_m(D) \quad . \tag{2.102}$$

But for $R_m(D) > C$, $R > C$ and therefore from the Converse Information Transmission Theorem (Berger [4]) the block code does not yield a distortion level less than D.

<u>**Q.E.D**</u>

As in Section 2.3, the results extend to the time average per-message distortion measure and the associated distortion function defined in (2.14). Corollary 2.2 establishes the information transmission theorem and its converse for this case.

Corollary 2.2

Theorems 2.3 and 2.4 remain valid, if the rate distortion function is replaced with the rate distortion function, $R_t(D)$, incorporating the time average per-message distortion measure (2.12).

Proof

Same technique as Corollary 2.1.

Q.E.D

The rate distortion function $R_m(D)$ is given an operational definition in terms of a communication system by Theorems 2.3 and 2.4. The theorems show that transmission of information using a message block code with a fidelity constraint D and with a finite average message delay constraint over a channel of capacity C is possible if $C > R_m(D)$ and is impossible if $R_m(D) \geq C$. If $R_m(D) \geq C$, then transmission can be achieved if one of the two constraints is removed.

## 2.5 Conclusions

This chapter discussed the performance bounds of source encoding under a fidelity criteria for sources found in communication networks. The source model of network message sources was presented.

The rate distortion function for this source was calculated and was shown to be equal to another rate distortion relationship. Source coding and converse source coding theorems were proved. These theorems provided an operational definition of the rate distortion function in terms of the existence of message block codes which satisfy the distortion criteria, and have a finite limiting average message delay associated with them. This chapter concluded with a discussion of the information transmission theorems for network message sources.

# CHAPTER III

## DELAY DISTORTION RELATIONSHIPS AND ADAPTIVE DATA COMPRESSION

The proofs of the information transmission theorems in Chapter II have indicated how to construct coding theorems which approach the performance predicted by the rate distortion function. These schemes have an arbitrary large message delay associated with them. Large message delays imply a structurally complex communication system (large buffer sizes required), and degraded service to the subscribers to the network. Therefore, for the design of actual networks, a more meaningful performance measure than the rate distortion function is the delay-rate-distortion relationship. The latter gives the minimum average message delay when constraints are placed on both the rate of transmission, and the fidelity of the reproduction of the source. For a given network configuration with a fixed rate of transmission, this relationship reduces to the delay distortion relationship which this chapter discusses. The network configuration and the delay distortion relationship, $\gamma(D)$, are introduced in Section 3.1. In Section 3.2, an adaptive data compression scheme is presented. This scheme utilizes a decision policy to minimize the average message delay for a given distortion level D, resulting in the delay distortion relationship $\gamma_A(D)$. The techniques from Markov decision theory that pertain to the analysis of the adaptive data compression scheme are presented in Section 3.3. Finally, using these techniques the structure of the optimum decision policy for a

single channel network is determined in Section 3.4. This network is an adequate model of a satellite channel or a radio relay channel. The technique developed here to determine the optimum policy can also be applied in the optimization of controls in a variety of other queueing and flow control problems found in networks.

## 3.1 The Delay Distortion Relationship and the System Model

In a communication network, messages which are transmitted to distant users accrue time delays. As the information contents of the messages increases, the demands on the network are heightened and the message delays are increased. For many message sources, such as analog waveforms (speech signal, telemetry waveform), to achieve an acceptable level of message delay for the transmission of a message through an existing network, the messages must be approximate prior to transmission. The messages are approximated by other messages, termed compressed messages, whose information contents are smaller than the information contents of the original messages. Since the compressed messages are approximations to the original messages, the fidelity of the reproduction of the messages is degraded. However, at the same time, the message delay has been reduced. As an example, consider the transmission of voice or other random analog waveforms through a packet-switching store-and-forward communications network, such as a computer communication network described in Kleinrock [10], or through a TDMA system such as a satellite communication network described in Abramson and Kuo [7]. From information theory, the entropy rate of such a source is unbounded, resulting in unbounded

message delays for any network realization if the messages are not approximated. Hence, an appropriate method of approximating or data compressing the messages is required, to insure that tolerable levels of message delay and fidelity are achieved.

To study the trade-off between delay and distortion, Rubin [13] introduced the delay distortion relationship, $\gamma(D)$. For a given source model, network, and routing procedure associated with the network, this relationship specifies the minimum message delay achieved when choosing a scheme among all data compression schemes belonging to some class of data compression procedures which achieve a fixed distortion level. In order to define $\gamma(D)$, let $T(dc)$ be the message delay for the given models, and for a data compression scheme dc. Furthermore, let $\widetilde{DC}$ be a class of data compression schemes. Then the delay distortion relationship is given by

$$\gamma(D) = \inf_{dc \in DC(D)} T(dc) \tag{3.1}$$

where

$$DC(D) = \{dc \in \widetilde{DC}: \text{ average distortion level associated with } dc \leq D\}. \tag{3.2}$$

The source model considered in the study of $\gamma(D)$ is the model described in Section 2.1 which emits randomly occurring messages. The renewal process which describes the message arrivals is assumed to be a Poisson process. Thus, the interarrival distribution function is given by

$$F_\tau(\tau) = 1 - e^{-\lambda\tau} \quad , \qquad \tau \geq 0 \ . \tag{3.3}$$

where $\lambda^{-1}$ is the mean interarrival time. This model of the message arrival process is used extensively in the analysis of communication networks and queueing systems. Further justification for the Poisson assumption are the limit theorems in Khinchine [34]. These theorems provide, under certain regularity conditions, that if the arrival processes of various sources are independent, then the superposition of the arrival processes approaches (as the number of processes increases and the individual process rate decreases) a Poisson process.

The fidelity of the reproduction of the source at the user is measured via the average per-message distortion measure given by

$$D = \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N} d_n \tag{3.4}$$

where $d_n$ is the average distortion associated with the $n^{th}$ message. So, $d_n$ is given by

$$d_n = E[\rho(x_n, y_n)] \tag{3.5}$$

where $\rho(\cdot,\cdot)$ is the per-message distortion measure described in Section 2.1, $x_n \in X$ is the $n^{th}$ message, and $y_n \in Y$ is the $n^{th}$ reproduction.

The network considered is an isolated path in a communication network which operates in a store-and-forward manner depicted in Figure 3.1. The path consists of M channels connected by nodes which store messages received in an incoming channel until the corresponding outgoing channel is free to retransmit the message. The messages are then transmitted on a first received, first transmitted basis. Each channel or branch, $b_m$, is assumed to transmit information bits about

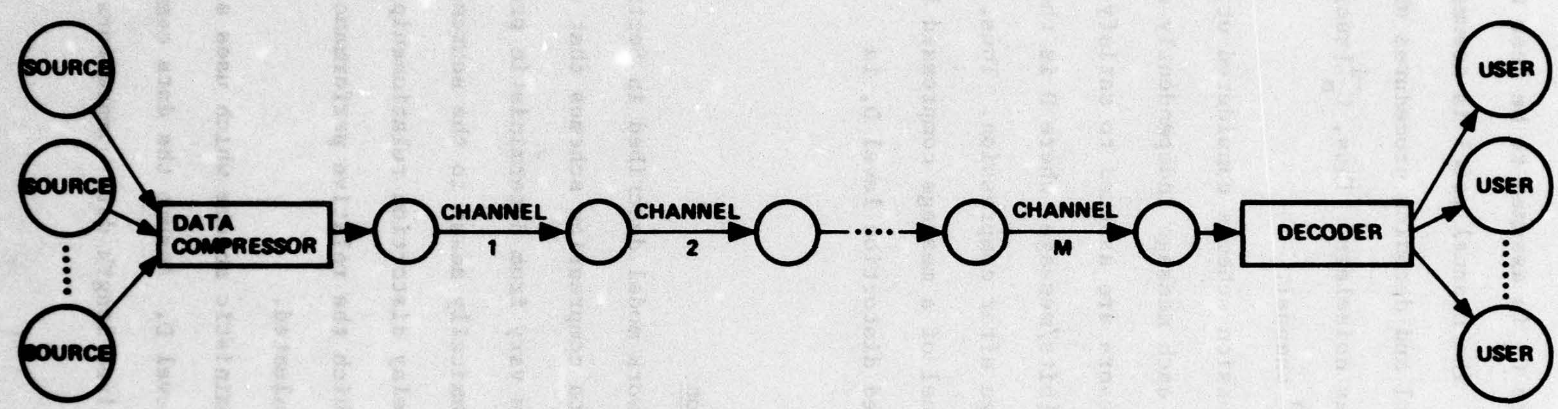



Figure 3.1. Data Compression for Multiple Sources in a Tandem Channel Communication Network

the source at a rate $C_m$, where $C_m$ is assumed to be less than the capacity of the $m^{th}$ channel. The channel model is assumed to incorporate the appropriate channel and decoding procedures which allow the channels to be regarded as noiseless. Thus, $C_m^{-1}$[sec/bit] is the transmission delay in the $m^{th}$ channel.

The class of data compression schemes considered utilizes data compressors which operate on each message independently of the other messages. These data compressors are assumed to satisfy a rate distortion relationship $\hat{r}(D)$[bits/message] where D is the average distortion accrued per-message after compression. Thus, the transmission time in the $m^{th}$ channel of a message compressed by a data compressor, with an associated distortion level D, is

$\hat{r}(D)\ C_m^{-1}$[sec/mess].

## 3.2 Adaptive Data Compression

For the source and network model described in Section 3.1, there are many classes of data compression schemes that warrant consideration. These schemes vary from deterministic processing schemes to schemes which dynamically adapt to the sources, and the state of the network. The delay distortion relationship, defined in Section 3.1, is a means by which the relative performance of various classes of schemes can be evaluated.

First, consider a deterministic scheme which uses a single data compressor with distortion level D. Assume the data compressor is selected from the class of fixed length data compressors which yield

a per-message rate distortion relationship $\hat{r}(D)$ where $\hat{r}(D)$ is described in Section 3.1. Then clearly, the message queueing delay is equivalent to the queueing delay of a customer in a M/D/1 tandem queueing system with Poisson arrivals of rate $\lambda$, and message service times in the $m^{th}$ channel, $m = 1, 2, \ldots, M$, equal to $\hat{r}(D)\, C_m^{-1}$. From Rubin [12] the limiting average message delay, $T_1(D)$, associated with this data compression scheme is given by the sum of the transmission delay and the queueing delay, and so,

$$T_1(D) = \begin{cases} \sum_{m=1}^{M} \hat{r}(D)\, \hat{C}_m^{-1} + \dfrac{\hat{r}(D) C_{min}^{-1}}{2} \dfrac{\lambda\, \hat{r}(D)}{C_{min} - \lambda\, \hat{r}(D)} & , \text{ if } D > D_{min}(\lambda) \\ \infty & , \text{ if } D \leq D_{min}(\lambda) \end{cases} \tag{3.6}$$

where

$$C_{min} = \min_{m \in [1,2,\ldots,M]} C_m \tag{3.7}$$

and

$$D_{min}(\lambda) = \hat{r}^{-1}(C_{min}/\lambda) \quad . \tag{3.8}$$

The delay distortion relationship, $\gamma_1^M(D)$, for such a scheme which uses a single data compressor is given by minimizing the limiting average message delay over all data compressors in the class of data compressors which yield a distortion level of D or less. So, from (3.1)

$$\gamma_1^M(D) = \inf_{\hat{D} \leq D} T_1(\hat{D}) \quad . \tag{3.9}$$

Then assuming $\hat{r}(D)$ to be a strictly decreasing function of D and using (3.6), the delay distortion relationship reduces to

$$\gamma_1^M(D) = T_1(D)$$

$$= \begin{cases} \sum_{m=1}^{M} \hat{r}(D)\, C_m^{-1} + \dfrac{\hat{r}(D)\, C_{min}^{-1}}{2} \dfrac{\lambda\, \hat{r}(D)}{C_{min} - \lambda\, \hat{r}(D)} & , \text{ if } D > D_{min}(\lambda) \\ \infty & , \text{ if } D \leq D_{min}(\lambda) \end{cases} \tag{3.10}$$

which is easily evaluated.

The scheme using a single data compressor can be extended by the introduction of K-1 additional fixed length data compressors. So, a message arriving proceeds immediately to one of the K data compressors. Following its compression, the message enters a buffer and awaits transmission in the network. The decision policy $\pi$ which directs the messages to the various data compressors is an element of a wide class of decision policies $\Pi$ which is specified later. The policy can be random or deterministic, stationary or time varying, dependent or independent of the past history. The major restriction placed upon this policy is that it be causal. Thus, it can not be dependent upon explicit knowledge of future events. Such a scheme is depicted in Figure 3.2.

For such a scheme to operate it is imperative that the decoder and user be informed which data compressor was used to operate on a given message. Without this knowledge, the performance of all but the simplest deterministic policies are considerably degraded. To

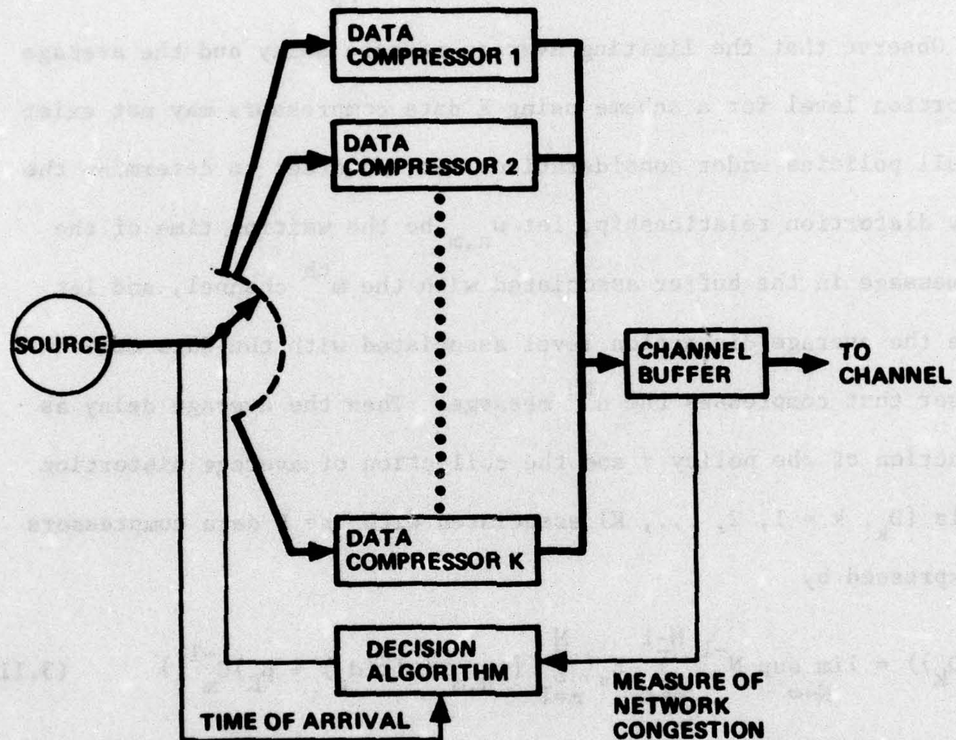



**Figure 3.2. Adaptive Data Compression Scheme Employing K Data Compressors**

provide this information, protocol information needs to be transmitted through the network. The protocol information used to inform the decoder is modelled as a fixed bit length segment $p_K$ which is attached to the bit length of the compressed messages. Hence, the total bit length of a message following compression by a compressor with associated distortion level D is $\hat{r}(D) + p_K$.

Observe that the limiting average message delay and the average distortion level for a scheme using K data compressors may not exist for all policies under consideration. So, in order to determine the delay distortion relationship, let $w_{n,m}$ be the waiting time of the $n^{th}$ message in the buffer associated with the $m^{th}$ channel, and let $d_n$ be the average distortion level associated with the data compressor that compresses the $n^{th}$ message. Then the average delay as a function of the policy $\pi$ and the collection of average distortion levels $\{D_k, k = 1, 2, \ldots, K\}$ associated with the K data compressors is expressed by

$$T_\pi^K(\{D_k\}) = \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_\pi\{ \sum_{m=1}^{M} [w_{n,m} + (\hat{r}(d_n) + p_K)C_m^{-1}]\} \qquad (3.11)$$

where all the buffers are initially assumed to be empty. This expression assumes that the total message arrives at the instant of arrival and that no delay is accrued from the compressing procedure itself. In a similar manner, the average distortion is given by

$$D_\pi^K(\{D_k\}) = \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_\pi(d_n) \quad . \qquad (3.12)$$

Then, the delay distortion relationship for schemes which use K data compressors is given by

$$\gamma_K^M(D) = \inf_{\substack{\pi\varepsilon\Pi(D;\{D_k\}) \\ \{D_k \geq 0, k=1,2,\ldots,K\}}} T_\pi^K(\{D_k\}) \tag{3.13}$$

where

$$\Pi(D;\ \{D_k\}) = \{\pi\ \varepsilon\ \Pi:\ \ D_\pi^K(\{D_k\}) \leq D\}\ .$$

Taking the minimum of $\gamma_K^M(D)$ over all possible values of K, the delay distortion relationship $\gamma_A^M(D)$ for adaptive schemes is expressed by

$$\gamma_A^M(D) = \min_{K\varepsilon[1,2,\ldots]} \gamma_K^M(D)\ . \tag{3.14}$$

There must be a finite K attaining the minimum in (3.14). This is due to the protocol information, $p_K$, being an increasing function of K and so, eventually, as K increases, the message delay in the network will be unbounded due to the excessive amount of protocol information. Thus, the optimal adaptive scheme uses a finite number of data compressors.

In order to compute $\gamma_A^M(D)$, the problem of finding $\gamma_K^M(D)$ must be considered first. Once $\gamma_K^M(D)$ is determined, $\gamma_A^M(D)$ is the lower envelope of the curves $\gamma_K^M(D)$, K = 1, 2, ... . To evaluate $\gamma_K^M(D)$, introduce the Lagrange multiplier $\mu \geq 0$ and define the function $L_\pi^K(\{D_k\})$ by

$$L_\pi^K(\{D_k\}) = T_\pi^K(\{D_k\}) + \mu\ D_\pi^K(\{D_k\}) \tag{3.15}$$

So, if the infimum operation in (3.13) can be replaced with the corresponding minimum operation (which is shown later to be valid), the expression for $\gamma_K^M(D)$ reduces to

$$\gamma_K^M(D^*) = \{ \min_{\substack{D_k \geq 0, \\ k=1,2,\ldots K}} \min_{\pi \varepsilon \Pi} L_\pi^K(\{D_k\})\} - \mu D^* \tag{3.16}$$

where

$$D^* = D_{\pi^*}^K(\{D_k^*\}) \tag{3.17}$$

and $\{D_k^* \geq 0\}$ and $\pi^* \varepsilon \Pi$ minimize $L_\pi^K(\{D_k\})$. Hence, (3.16) and (3.17) describe parameterically the delay distortion relationship as a function of the Lagrange multiplier $\mu$.

To determine when (3.16) is valid, it is necessary to investigate $L^K(\{D_k\})$, given by

$$L^K(\{D_k\}) = \inf_{\pi \varepsilon \Pi} L_\pi^K(\{D_k\}) \quad , \tag{3.18}$$

and ascertain if there exists a $\pi^* \varepsilon \Pi$ which satisfies

$$L^K(\{D_k\}) = L_{\pi^*}^K(\{D_k\}) \quad . \tag{3.19}$$

This investigation can be achieved by considering the following problem, where the network is restricted to a single channel with transmission rate C. The tandem channel case is treated in Chapter V.

<u>Single Channel Adaptive Data Compression Problem</u> (SCADCP)

Consider a data compression system, as described in this section, which uses K data compressors whose associated distortion levels are $D_k$, k = 1, 2, ..., K. The communication network is a single channel of transmission rate C. The state of the system is observed at times of arrivals of messages. When a message arrives it must be assigned

to one of the K possible data compressors for compression. Let the state of the communication system at the instant of the $n^{th}$ message arrival, n = 0, 1, 2, ..., be defined as $x_n$. The state $x_n$ is defined as the vector whose components are the waiting time, $w_n$, of the $n^{th}$ message, and the interarrival time, $\tau_n$, between the n-1 message and the $n^{th}$ message arrival times. Thus,

$$x_n = (w_n, \tau_n)$$

and the state space is given by $X = ([0, \infty) \times [0, \infty))$. The decision at the $n^{th}$ arrival is $a_n \in [1, 2, \ldots, K]$, where $a_n$ indicates the compressor that operates on the $n^{th}$ message. The history $H_n$ of the communication system up to the $n^{th}$ arrival is given by

$$H_n = (x_0, a_0, x_1, a_1, \ldots, x_n, a_n) \quad . \tag{3.20}$$

Clearly, $H_n$ is a sufficient exhaustive description of the history of the communication system, since the communication system can be considered to be a single server queueing system, and the past history of such a queueing system is specified by the past interarrival times and the past service times. From the recurrence relationship for the waiting time of a queueing system, $w_{n+1}$ is given by

$$w_{n+1} = \max(w_n + t_{a_n} - \tau_{n+1}, 0) \tag{3.21}$$

where the transmission time, $t_k$, associated with a message compressed by the $k^{th}$ data compressor is given by, $k \in [1, 2, \ldots, K]$,

$$t_k = (r(D_k) + p_K)C^{-1} \quad . \tag{3.22}$$

So, using (3.19) a version of the transition probability distribution, $\Pr(w_{n+1} \leq W, \tau_{n+1} \leq T | H_n)$, is given by

$$\Pr(w_{n+1} \leq W, \tau_{n+1} \leq T | H_n) = \Pr(w_{n+1} \leq W, \tau_{n+1} \leq T | (w_n, \tau_n), a_n)$$

$$= \begin{cases} F_\tau(T) & , \text{if } W \geq w_n + t_{a_n} \\ F_\tau(T) - F_\tau(w_n + t_{a_n} - W) & , \text{if } T \geq w_n + t_{a_n} - W > 0 \\ 0 & , \text{otherwise} \end{cases} \tag{3.23}$$

A policy $\pi$ for controlling the system is described by a set of functions $\{Q_k^n(H_{n-1}, x_n), k = 1, 2, \ldots, K, n \geq 0\}$. The functions $Q_k^n(\cdot)$ are measurable with respect to the sigma algebra generated by $H_{n-1}$ and $x_n$ and the functions satisfy for k = 1, 2, ..., K

$$Q_k^n(H_{n-1}, x_n) \geq 0 \qquad \text{w.p.1} \tag{3.24}$$

and

$$\sum_{k=1}^{K} Q_k^n(H_{n-1}, x_n) = 1 \qquad \text{w.p.1} \quad . \tag{3.25}$$

These functions represent the probabilities of choosing the various data compressors given the past history and the present state. Thus, $Q_k^n(H_{n-1}, x_n)$ is the probability of operating on the $n^{th}$ message with the $k^{th}$ data compressor given history $H_{n-1}$ and state $x_n$. Denote by $\Pi$ the set of all such policies.

To measure the effectiveness of a rule $\pi \in \Pi$, a cost structure is introduced. If the system is in state $x \in X$ at the $n^{th}$ arrival and action a $\in$ [1, 2, ..., K] is taken, a cost c(x, a) is incurred.

This cost is given by

$$c(x,a) = w + t_a + \mu D_a \tag{3.26}$$

where $x = (w, \tau)$. Thus, the average cost $\Phi(x, \pi)$ under rule $\pi \in \Pi$, when the system is initially in state $x \in X$, is expressed by

$$\Phi(x,\pi) = \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_\pi\{c(x_n,a_n)|x_0 = x\} . \tag{3.27}$$

and

$$\Phi(x) = \inf_{\pi\in\Pi} \Phi(x, \pi) . \tag{3.28}$$

Clearly from (3.11), (3.12) and (3.15)

$$L_\pi^K(\{D_k\}) = \Phi((0,0), \pi) \tag{3.29}$$

where $L_\pi^K(\{D_k\})$ is computed assuming the system is originally free of messages, and then from (3.18)

$$L^K(\{D_k\}) = \Phi((0,0)) . \tag{3.30}$$

The problem is to determine when a policy $\pi^* \in \Pi$ exists such that for all $x \in X$

$$\Phi(x) = \Phi(x, \pi^*) \tag{3.31}$$

The existence of such a policy is established in Section 3.4 by noting that the process $\{(x_n, a_n), n = 0, 1, 2, \ldots\}$ is a Markov decision process as described in Section 3.3. From (3.29) and (3.30) the determination of a $\pi^*$ that satisfies (3.31) implies, the existence of a $\pi^*$ which satisfies (3.19). The structure of the

policy $\pi^*$ that satisfies (3.19) is determined in Section 3.4.

## 3.3 Markov Decision Processes

In this section Markov decision processes are studied with the emphasis of solving the average cost criteria problem for the case of unbounded cost functions and nondenumerable state spaces. A Markov decision process, (see, for example Ross [19]) is specified by four quantities: A state space, an action space, a law of motion, and a reward or cost function. Assume that the state space X is a complete separable metric space and that the action space is a finite set A. At time n = 0, 1, 2, ..., the state of the process under consideration is $x_n \in X$ and action $a_n \in A$ is selected. Then the history $H_n$ of the process is denoted by the sequence of states and actions,

$$H_n = \{x_0, a_0, x_1, a_1, \ldots, x_n, a_n\} \quad . \tag{3.32}$$

To specify the law of motion let $\mathcal{B}$ be the σ-algebra of Borel subsets of X. Then assume that for every $x \in X$ and $a \in A$, there is a known probability measure $Pr(\cdot|x,a)$ on $\mathcal{B}$ such that, for some version of the measure,

$$Pr(x_{n+1} \in B|x_n = x, a_n = a, H_{n-1}) = Pr(B|x, a) \tag{3.33}$$

for every $B \in \mathcal{B}$ and all histories $H_{n-1}$.

The actions are selected according to some conditional probability distribution

$$\Pr(a_n = a | H_{n-1}, x_n) = Q_a^n(H_{n-1}, x_n) \quad . \tag{3.34}$$

The functions $Q_a^n(H_{n-1}, x_n)$ are measurable with respect to the $\sigma$-algebra generated by the random variables $H_{n-1}$ and $x_n$. These functions satisfy

$$Q_a^n(H_{n-1}, x_n) \geq 0 \quad \text{w.p.1} \tag{3.35}$$

and

$$\sum_{a \varepsilon A} Q_a^n(H_{n-1}, x_n) = 1 \quad \text{w.p.1} \quad . \tag{3.36}$$

Denoted by $\Pi$ the set of all such policies which generate actions in this manner. A policy $\pi \varepsilon \Pi$ is said to be stationary if for all $a \varepsilon A$ and $n \geq 0$

$$Q_a^n(H_{n-1}, x_n) = Q_a(x_n) \quad \text{w.p.1} \tag{3.37}$$

and stationary deterministic if $Q_a(x_n)$ equals 0 or 1 w.p.1 for all $a \varepsilon A$ and $n \geq 0$.

The reward function is specified by the cost function $c(x,a)$ which is the cost incurred when the process is observed in state $x \varepsilon X$ and action $a \varepsilon A$ is taken. The cost function is assumed to be finite, but not necessarily bounded.

The process $\{(x_n, a_n), n = 0, 1, 2, \ldots\}$ is termed a Markov decision process. Two possible measures of performance of a policy $\pi \varepsilon \Pi$ that governs the process are the expected total $\alpha$-discounted cost, $V_\alpha(x,\pi)$, given for $\alpha \varepsilon (0,1)$ by

$$V_\alpha(x,\pi) = \limsup_{N\to\infty} \sum_{n=0}^{N-1} \alpha^n E_\pi\{c(x_n,a_n)|x_0 = x\} \tag{3.38}$$

and the average cost, $\Phi(x,\pi)$, given by

$$\Phi(x,\pi) = \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_\pi\{c(x_n,a_n)|x_0 = x\} . \tag{3.39}$$

Minimizing over $\pi \in \Pi$ in (3.38) and (3.39), let

$$V_\alpha(x) = \inf_{\pi\in\Pi} V_\alpha(x,\pi) \tag{3.40}$$

and

$$\Phi(x) = \inf_{\pi\in\Pi} \Phi(x,\pi) \quad . \tag{3.41}$$

Results pertaining to Markov decision processes deal with establishing when a stationary deterministic policy $\pi^* \in \Pi$ minimizes $V_\alpha(x,\pi)$ or minimizes $\Phi(x,\pi)$. The development that follows discusses the results for the average cost criteria for processes with unbound cost functions and nondenumerable state spaces. The results are an extension of Lippman's [21] results.

To determine when a stationary deterministic optimal rule exists for the average cost criteria, the usual technique is to find the limit of the policies which solve the $\alpha$-discounted cost problem as $\alpha$ approaches 1. Then ascertain if this limit is the optimal stationary deterministic policy for the average cost problem. First, an assumption needs to be made to insure that $V_\alpha(x,\pi)$ exists and is finite for each $\alpha \in (0,1)$, $x \in X$, and $\pi \in \Pi$.

**Assumption 3.1**

There exists an integer $m \geq 1$, a real valued function $g(\cdot)$ on X with $g(x) \geq 1$ for all $x \varepsilon X$ and a real number $b \geq 0$ such that

$$L \equiv \sup_{x \varepsilon X} \{|\max_{a \varepsilon A} c(x,a)|g(x)^{-m}\} < \infty \quad , \tag{3.42}$$

and for all $x \varepsilon X$ and for $n = 1, 2, \ldots, m$,

$$\max_{a \varepsilon A} \int_X g(\hat{x})^n \, P(d\hat{x}|x,a) \leq (g(x) + b)^n \tag{3.43}$$

Using Assumption 3.1, Lippman [21] proves a theorem which guarantees the existence of an optimal stationary deterministic policy for the $\alpha$-discounted cost problem. The relevent results of this theorem that are required for the development of the solution to the average cost problem are stated in the following theorem.

**Theorem 3.1**

For a Markov decision process, suppose Assumption 3.1 holds, then an optimal stationary deterministic policy $\pi_\alpha^*$ exists for the $\alpha$-discounted cost problem, $\alpha \varepsilon (0,1)$, and $V_\alpha(x,\pi_\alpha^*)$ is the unique solution to

$$V_\alpha(x,\pi_\alpha^*) = \min_{a \varepsilon A} \{c(x,a) + \alpha \int_X V_\alpha(\hat{x},\pi_\alpha^*) \, Pr(d\hat{x}|x,a)\} \quad , \tag{3.44}$$

the functional equation of dynamic programming. In addition, $\pi_\alpha^*$ is the policy which selects an action minimizing the right side of (3.44) for each $x \varepsilon X$.

**Proof**

See Lippman [21] for proof.

Q.E.D

A consequence of the proof of the theorem is a policy improvement algorithm for finding $V_\alpha(x,\pi_\alpha^*)$, which is stated in a corollary.

**Corollary 3.1**

Suppose Assumption 3.1 is valid, and let $\alpha \in (0,1)$. Let $U_{0,\alpha}(x) = 0$ for all $x \in X$ and

$$U_{n+1,\alpha}(x) = \min_{a \in A} \{c(x,a) + \alpha \int_X U_{n,\alpha}(\hat{x})\, Pr(d\hat{x}|x,a)\} \ , \qquad (3.45)$$

then

$$V_\alpha(x,\pi_\alpha^*) = \lim_{n\to\infty} U_{n,\alpha}(x) \qquad (3.46)$$

**Proof**

A consequence of the proof of Theorem 1 in Lippman [21].

Q.E.D

Before proceeding to the average cost problem, a lemma is presented which is needed to show the relationship between the limit of the $\alpha$-discounted cost and the average cost for a policy $\pi \in \Pi$. This lemma is an extension of an Abelian theorem in Widder [35].

Lemma 3.1

Suppose Assumption 3.1 is valid, then

$$\Phi(x,\pi) \geq \limsup_{\alpha \uparrow 1} (1-\alpha) V_\alpha(x,\pi) \tag{3.47}$$

for all $x \in X$ and $\pi \in \Pi$.

Proof

Let $\alpha \in (0,1)$, and let

$$W_\alpha(x,\pi) = \limsup_{N\to\infty} E_\pi\{ \sum_{n=0}^{N-1} \alpha^n \sum_{j=0}^{n} c(x_j,a_j) | x_0 = x\}$$

$$= \limsup_{N\to\infty} E_\pi\{[ \sum_{n=0}^{N-1} \frac{\alpha^n}{1-\alpha} c(x_n,a_n) - \frac{\alpha^N}{1-\alpha} \sum_{n=0}^{N-1} c(x_n,a_n)] | x_0 = x\} \tag{3.48}$$

for all $x \in X$, and $\pi \in \Pi$. But from Assumption 3.1 for all $x \in X$ and $a \in A$

$$c(x,a) \leq L\, g(x)^m \tag{3.49}$$

and from (3.43) for all $\pi \in \Pi$, $n > 0$, and $x \in X$, the bound

$$E_\pi\{c(x_n,a_n) | x_0 = x] \leq (g(x) + (n-1)b)^m \tag{3.50}$$

is obtained. Thus, one obtains

$$\limsup_{N\to\infty} \frac{\alpha^N}{1-\alpha} E_\pi \sum_{n=0}^{N-1} c(x_n;a_n) | x_0 = x) = 0 \ . \tag{3.51}$$

So, substituting (3.38) and (3.51) into (3.47),

$$W_\alpha(x,\pi) = \limsup_{N\to\infty} E_\pi\{\sum_{n=0}^{N-1} \frac{\alpha^n}{1-\alpha} c(x_n,a_n)|x_0 = x\}$$

$$= \frac{V_\alpha(x,\pi)}{1-\alpha} \tag{3.52}$$

Rewriting (3.52), for all $M \geq 0$

$$(1-\alpha)\, V_\alpha(x,\pi) = W_\alpha(x,\pi)\,(1-\alpha)^2$$

$$= (1-\alpha)^2 \sum_{n=0}^{M} \alpha^n \sum_{j=0}^{n} E_\pi\{c(x_j,a_j)|x_0 = x\}$$

$$+ \limsup_{N\to\infty} (1-\alpha)^2 \sum_{n=M+1}^{N-1} n\alpha^n \sum_{j=0}^{n} E_\pi\{c(x_j,a_j)|x_0 = x\}/n \tag{3.53}$$

Now consider the bounds which state that for all $m = M+1, \ldots, N-1$

$$\frac{1}{m}\sum_{j=0}^{m} E_\pi\{c(x_j,a_j)|x_0 = x\} \leq \sup_{N>m'>M} \frac{1}{m'} \sum_{j=0}^{m'} E_\pi\{c(x_j,a_j)|x_0 = x\} \tag{3.54}$$

and

$$\sum_{n=M+1}^{N-1} n\alpha^n \leq \frac{\alpha}{(1-\alpha)^2} \quad . \tag{3.55}$$

Using these bounds in (3.53), results in

$$(1-\alpha)\, V_\alpha(x,\pi) \leq (1-\alpha)^2 \sum_{n=0}^{M} \alpha^n \sum_{j=0}^{n} E_\pi\{c(x_j,a_j)|x_0 = x\}$$

$$+ \limsup_{N\to\infty} \sup_{N>m>M} \alpha\{\sum_{j=0}^{m} E_\pi\{c(x_j,a_j)|x_0 = x\}/m\} \tag{3.56}$$

Taking the lim sup of (3.56) as $\alpha \uparrow 1$ and noting that for $0 \leq n \leq M$

$$\sum_{j=0}^{n} E_\pi(c(x_j,a_j)|x_0 = x) < \infty \quad , \tag{3.57}$$

it is clear that

$$\limsup_{\alpha \uparrow 1} (1-\alpha)\, V_\alpha(x,\pi) \leq \sup_{m \geq M+1} \{ \sum_{j=0}^{m} E_\pi[c(x_j,a_j)|x_0 = x]/m\} \tag{3.58}$$

Thus, letting M become arbitrarily large, and using (3.39), the result, (3.47), is achieved.

Q.E.D

Further assumptions need to be made to guarantee that some stationary deterministic policy is average optimal. These assumptions deal with the convergence of the $\alpha$-discounted minimal cost function to the minimal average cost function, and differ from the assumptions made in Lippman [21].

Assumption 3.2

There exists a state $x^* \varepsilon X$, a value $\alpha^* \varepsilon (0,1)$, and a real valued function $L(x)$ on $X$ such that for all $\alpha \varepsilon (\alpha^*,1)$ and $x \varepsilon X$

$$|V_\alpha(x^*,\pi^*_\alpha) - V_\alpha(x,\pi^*_\alpha)| \leq L(x) \tag{3.59}$$

when $\pi^*_\alpha$ exists.

Assumption 3.3

There exists a stationary policy $\hat{\pi}$, and an increasing sequence $\{\alpha_n\}$ with $\alpha_n \uparrow 1$ that satisfies

$$\Phi(x,\hat{\pi}) = \limsup_{n\to\infty} (1-\alpha_n)\, V_\alpha(x,\pi^*_{\alpha_n}) \tag{3.60}$$

for all $x \varepsilon X$.

A technique for finding $\hat{\pi}$ in Assumption 3.3 is to place a further assumption on the problem to force $\hat{\pi}$ to be the limiting policy of $\{\pi^*_{a_n}\}$. For the case of a continuous state space, the limit of a set of policies is a difficult concept to define. In verifying the assumptions for the SCADCP outlined in Section 3.2, the limit of the policies $\pi^*_{\alpha_n}$ for the problem can be defined quite easily. Hence, $\hat{\pi}$, which is the limit of the policies, can be found. Section 3.4 discusses these results.

The following theorem proves that if a Markov decision process satisfies the Assumptions 3.1, 3.2 and 3.3, then an average optimal stationary deterministic policy exists.

<u>Theorem 3.2</u>

Suppose Assumptions 3.1, 3.2, and 3.3 hold, then the average optimal policy $\pi^*$ is a stationary deterministic policy given by the policy $\hat{\pi}$ in Assumption 3.3 and for all $x \in X$, $\Phi(x)$ in (3.41) is independent of $x$.

<u>Proof</u>

From Assumption 3.1 and Theorem 3.1, for all $\alpha \in (0,1)$, $\pi^*_\alpha$ exists and is a stationary deterministic policy. From Lemma 3.1 for any increasing sequence $\{\alpha_n\}$ with $\alpha_n \uparrow 1$,

$$\Phi(x,\pi) \geq \limsup_{n\to\infty} (1-\alpha_n)\, V_{\alpha_n}(x,\pi) \qquad (3.61)$$

for all $x \in X$ and all $\pi \in \Pi$. Since $\Pi^*_\alpha$ is the $\alpha$-optimal discounted cost policy, then for all $\alpha \in (0,1)$

$$(1-\alpha)\ V_\alpha(x,\pi) \geq (1-\alpha)\ V_\alpha(x,\pi^*_\alpha)$$

Taking the lim sup with respect to $\alpha$, and using (3.60) and Assumption 3.2,

$$\Phi(x,\pi) \geq \limsup_{n\to\infty} (1-\alpha_n)\ V_{\alpha_n}(x,\pi^*_{\alpha_n}) = \limsup_{n\to\infty} (1-\alpha_n)\ V_{\alpha_n}(x^*,\pi^*_{\alpha_n}). \tag{3.62}$$

Finally, since the sequence $\{\alpha_n\}$ is arbitrary, applying Assumption 3.3 to (3.62) gives

$$\Phi(x,\pi) \geq \Phi(x,\hat{\pi}) = \Phi(x^*,\hat{\pi}) \tag{3.63}$$

for all $\pi \in \Pi$, and $x \in X$. Consequently, since (3.63) holds for all $x \in X$ and $\pi \in \Pi$, then

$$\Phi(x) = \Phi(x^*,\hat{\pi}) \tag{3.64}$$

and

$$\pi^* = \hat{\pi}\ . \tag{3.65}$$

<u>Q.E.D</u>

The procedures utilized in Theorem 3.2 differ from the procedures used in the classical works, as in Derman [17] and Ross [19]. In these works, first a theorem is proved stating that if the functional equation, which is similar in nature to (3.44), for the average cost problem has a solution, an average optimal stationary deterministic policy exists. Then to show the existence of such a solution, the limit of the functional equation (3.41) for the $\alpha$-discounted cost

problem is examined. This technique requires additional assumptions beyond Assumption 3.1 - 3.3 to guarantee the convergence of the functional equations. These additional assumptions might not be easily verified for Markov decision processes with unbounded cost functions and nondenumerable state spaces.

## 3.4 The Structure of Optimal Policy for the Single Channel Adaptive Data Compression Problem

The Single Channel Adaptive Data Compression Problem (SCADCP) outlined in Section 3.2 deals with the issue of optimally controlling the selection of data compressors. This issue can be categorized as a problem dealing with the optimal control of a queueing system. The techniques derived from examining Markov decision processes are used extensive in solving this class of problems. In this section, for the SCADCP where the adaptive compression scheme uses two data compressors (K=2), the structure of the optimal policy is determined using the results of Markov decision theory.

In order to apply the results in Section 3.3 to the SCADCP, it must be verified that a Markov decision process is being generated by the communication system under consideration. Clearly, the process $\{(x_n,a_n), n = 0, 1, 2, \ldots\}$ of the SCADCP is a Markov decision process since the state space $X = ([0,\infty) \times [0,\infty))$, action space $[1,2]$, the law of motion (3.23) and the cost function (3.26) all satisfy the requirements outlined in Section 3.3 for a Markov decision process. Thus, the structure of the optimal policy can be found by using Theorem 3.2, if Assumptions 3.1, 3.2 and 3.3 can be verified.

Before proceeding with the verification of the assumptions, assume without loss of generality for the remainder of the chapter that the distortion levels associated with the two data compressors are ordered in the following manner:

$$D_1 \leq D_2 \quad . \tag{3.66}$$

Furthermore, consider the two regions,

Region 1: $D_2 \leq \hat{r}^{-1}(\frac{C}{\lambda} - p_2)$

Region 2: $D_2 > \hat{r}^{-1}(\frac{C}{\lambda} - p_2)$ .

For region 1, for every $\pi \in \Pi$, $\Phi(x,\pi)$ is unbounded as shown in the following lemma.

Lemma 3.2

If

$$D_2 \leq \hat{r}^{-1}(\frac{C}{\lambda} - p_2) \quad , \tag{3.67}$$

then $\Phi(x,\pi)$ is unbounded for every $\pi \in \Pi$.

Proof

From (3.22) and (3.67), for $k = 1, 2$

$$t_k = (\hat{r}(D_k) + p_2)C^{-1} \geq \lambda^{-1} \tag{3.68}$$

and from (3.66),

$$t_1 \geq t_2 \quad . \tag{3.69}$$

Expression (3.21) for the waiting time gives

$$w_{n+1} = \max(0, w_n + t_{a_n} - \tau_{n+1}) \quad . \tag{3.70}$$

Thus, the policy $\hat{\pi}$ which minimizes $w_n$ for all $n \geq 0$ is the policy that always selects data compressor 2. So, for all $n \geq 0$

$$E_\pi(w_n | x = x_0) \geq E_{\hat{\pi}}(w_n | x = x_0) \quad . \tag{3.71}$$

Under $\hat{\pi}$, $w_n$ is the waiting time of the $n^{th}$ customer in a M/D/1 queueing system with service time $t_2$, and mean interarrival time $\lambda^{-1}$. From (3.68) the traffic intensity $\rho$ is given by

$$\rho = t_2/\lambda \geq 1 \tag{3.72}$$

and therefore by Lindley's theorem (Lindley [33]), $E_{\hat{\pi}}(w_n | x_0 = x)$ grows without bound (i.e., $\lim_{n\to\infty} E_\pi(w_n | x_0 = x) = \infty$). Now using (3.24) and (3.25), $\Phi(x,\pi)$ is bounded by

$$\Phi(x,\pi) \geq \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_\pi(w_n | x_0 = x)$$

$$\geq \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_{\hat{\pi}}(w_n | x_0 = x) \tag{3.73}$$

Thus, $\Phi(x,\pi)$ is unbounded.

<u>Q.E.D</u>

Consequently, the region of interest is

$$D_2 > r^{-1}(\frac{C}{\lambda} - p_2) \tag{3.74}$$

and the following set of lemmas verify the assumptions stated in

Section 3.3.

Lemma 3.3

Assumption 3.1 is valid for the SCADCP.

Proof

Let

$$m = 1 \quad ,$$

$$g((w,\tau)) = w + 1 \quad ,$$

and

$$b = \max_{a\varepsilon[1,2]} t_a < \infty$$

Then

$$\sup_{\substack{w\varepsilon[0,\infty) \\ \tau\varepsilon[0,\infty)}} \{ | \max_{a\varepsilon 1,2} c((w,\tau),a) | g((w,\tau))^{-m} \}$$

$$= \sup_{\substack{w\varepsilon[0,\infty) \\ \tau\varepsilon[0,\infty)}} \{ | \max_{a\varepsilon[1,2]} (w + t_a + \mu D_a) | (w+1)^{-1} \}$$

$$< 1 + \max_{a\varepsilon[1,2]} (t_a + \mu D_a) < \infty \qquad (3.75)$$

and (3.42) is verified.

To verify (3.43), using (3.23)

$$\max_{a\varepsilon[1,2]} \int_{\substack{\hat{w}\varepsilon[0,\infty) \\ \hat{\tau}\varepsilon[0,\infty)}} g((\hat{w},\hat{\tau}))\ Pr(d(\hat{w},\hat{\tau})\,|\,(w,\tau),a)$$

$$\leq \max_{a\varepsilon[1,2]} (w + 1 + t_a) \int_{\substack{\hat{w}\varepsilon[0,\infty) \\ \hat{\tau}\varepsilon[0,\infty)}} Pr(d(\hat{w},\hat{\tau})\,|\,(w,\tau),a)$$

$$\leq g((w,\tau)) + b \quad . \tag{3.76}$$

Hence, Assumption 3.1 is valid.

Q.E.D

By Theorem 3.1, there exists an $\alpha$-optimal stationary deterministic policy for all $\alpha \varepsilon (0,1)$ which can be found using the policy improvement algorithm of Corollary 3.1. The following two lemmas establish the structure of the $\alpha$-optimal policy.

Lemma 3.4

An $\alpha$-optimal policy exists for all $\alpha \varepsilon (0,1)$ and is a stationary deterministic policy that depends only on the observed waiting time. Hence, the $\alpha$-optimal policy is not a function of the observed interarrival time given the observed waiting time.

Proof

Lemma 3.3 and Theorem 3.1 imply that an $\alpha$-optimal policy exists and is a stationary deterministic policy. To show that the $\alpha$-optimal policy is independent of the observed interarrival time, first note

that from (3.26) $c((w,\tau),a)$ can be written $c(w,a)$ since it is not a function of $\tau$. Consider the policy improvement algorithm (3.44) where clearly $U_{0,\alpha}((w,\tau))$, given by

$$U_{0,\alpha}((w,\tau)) = 0 \tag{3.77}$$

for all $w \in [0,\infty)$ and $\tau \in [0,\infty)$, is independent of $\tau$. Now suppose $U_{n,\alpha}((w,\tau))$ is only a function of $w$ (i.e.; $U_{n,\alpha}((w,\tau)) = U_{n,\alpha}(w)$ for all $w \in [0,\infty)$ and $\tau \in [0,\infty)$). Then, it is easy to verify that $U_{n+1,\alpha}((w,\tau))$ as given by

$$U_{n+1,\alpha}((w,\tau)) = \min_{a\in[1,2]} \{c(w,a) + \alpha \int_{\substack{\hat{w}\in[0,\infty) \\ \hat{\tau}\in[0,\infty)}} U_{n,\alpha}((\hat{w},\hat{\tau}))\, Pr(d(\hat{w},\hat{\tau})\,|\,(w,\tau),a)\} \tag{3.78}$$

is not a function of $\tau$, since the conditional distribution $Pr(w_{n+1} \leq W,\ \tau_{n+1} \leq T\,|\,(w_n,\tau_n),a_n)$ in (3.23) is independent of $\tau_n$. Thus, by induction $U_{n,\alpha}((w,\tau))$ is independent of $\tau$ for all $n \geq 0$. By (3.46) $V_\alpha((w,\tau),\pi^*_\alpha)$ is given by the limit of $U_{n,\alpha}((w,\tau))$ as $n \to \infty$, and hence $V_\alpha((w,\tau),\pi^*_\alpha)$ is independent of $\tau$. Since the policy $\pi^*_\alpha$ minimizes the right side (3.44), and the right side of (3.44) is not a function of $\tau$, then $\pi^*_\alpha$ is a policy which, given the observed waiting time value, is independent of the corresponding observed interrarival time.

<u>Q.E.D</u>

Before proceeding, the definition of one type of stationary

deterministic policy is required. A stationary deterministic policy $\pi$ is said to be a connected policy if the state space X can be subdivided into disjoint connected regions $X_a$, a = 1, 2, ..., K, such that $\pi$ specifies action a is to be taken if, and only if, the system is observed in state $x \varepsilon X_a$. The following lemma establishes that the $\alpha$-optimal policy is a connected policy.

**Lemma 3.5**

For two data compressors, A = [1,2], the $\alpha$-optimal policy, $\alpha \varepsilon (0,1)$, is a connected policy. This policy specifies that data compressor one is selected when the waiting time w of a message satisfies $w \le T_\alpha$ and otherwise data compressor two is selected. $T_\alpha$ is termed the threshold and is selected to minimize the $\alpha$-discounted cost problem.

**Proof**

From Lemma 3.4 any dependencies on the interarrival times can be disregarded. So, substituting the value of $c((w,\tau),a)$, equation (3.24), and the expression for $Pr(w_{n+1} \le W, \tau_{n+1} \le T | (w_n,\tau_n),a_n)$, equation (3.23), the policy improvement algorithm (3.45) becomes

$$U_{n+1,\alpha}(w) = \min_{a\varepsilon[1,2]} \Big\{ w + t_a + \mu D_a + \alpha\Big[\int_0^{w+t_a} U_{n,\alpha}(w + t_a - \hat{w})F_\tau(d\hat{w}) + U_{n,\alpha}(0)\ (1 - F_\tau(w + t_a))\Big]\Big\} \qquad (3.79)$$

where

$$U_0(w) = 0 \qquad (3.80)$$

for $w \in [0,\infty)$. Now define

$$\hat{U}_{n,\alpha}(w) = \int_0^w U_{n,\alpha}(w-\hat{w})\ F_\tau(d\hat{w}) + U_{n,\alpha}(0)\ (1 - F_\tau(w)) \tag{3.81}$$

and so (3.75) becomes

$$U_{n+1,\alpha}(w) = \min_{a\varepsilon[1,2]} \{w + t_a + \mu D_a + \alpha\ \hat{U}_{n,\alpha}(w + t_a)\}\ . \tag{3.82}$$

In order to find $\pi^*_\alpha$ it is necessary to find the behavior of $U_{n,\alpha}(w)$ that satisfies (3.82). First, it is required to show that $U_{n,\alpha}(w)$ is a nondecreasing function of $w$ for all $n \geq 0$, and for all $\alpha \varepsilon (0,1)$. Clearly,

$$U_{1,\alpha}(w) = w + \min_{a\varepsilon[1,2]} (t_a + \mu D_a) \tag{3.83}$$

and so $U_{1,\alpha}(w)$ is a nondecreasing function of $w$, for $w \varepsilon [0,\infty)$. Now suppose $U_{n,\alpha}(w)$ is a nondecreasing function of $w$. Then from (3.81) $\hat{U}_{n,\alpha}(w)$ is a nondecreasing function and so $t_a + \mu D_a + \alpha\hat{U}_{n,\alpha}(w+t_a)$, $a = 1,2$, are nondecreasing functions. Thus, the lower envelope of $\{w + t_a + \mu D_a + \alpha\ \hat{U}_{n,\alpha}(w + t_a),\ a = 1, 2\}$, which is $U_{n+1,\alpha}(w)$, is a nondecreasing function of $w$. Therefore, by induction, $U_{n,\alpha}(w)$ is a nondecreasing function of $w$ for all $n \geq 0$ and $\alpha \varepsilon (0,1)$.

Now to show that the $\alpha$-optimal policy is a connected policy with threshold $T_\alpha$, define for $w < 0$

$$U_{n,\alpha}(w) = U_{n,\alpha}(0) \tag{3.84}$$

and

$$\hat{U}_{n,\alpha}(w) = \hat{U}_{n,\alpha}(0) \ . \tag{3.85}$$

Furthermore, define for $w \in (-\infty,\infty)$

$$\Delta U_{n,\alpha}(w) = U_{n,\alpha}(w + t_1) - U_{n,\alpha}(w + t_2) \tag{3.86}$$

and

$$\Delta\hat{U}_{n,\alpha}(w) = \hat{U}_{n,\alpha}(w + t_1) - \hat{U}_{n,\alpha}(w + t_2) \ . \tag{3.87}$$

By (3.81) $\Delta\hat{U}_{n,\alpha}(w)$ satisfies

$$\Delta\hat{U}_{n,\alpha}(w) = \int_0^\infty [U_{n,\alpha}(w + t_1 - \hat{w}) - U_{n,\alpha}(w + t_2 - \hat{w})]F_\tau(d\hat{w})$$

$$= \int_0^\infty \Delta U_{n,\alpha}(w - \hat{w})F_\tau(d\hat{w}) \ . \tag{3.88}$$

Clearly from (3.82), at stage n+1 the optimal policy is to select data compressor 1 if the waiting time w satisfies

$$t_1 - t_2 + \mu(D_1 - D_2) + \alpha\, \Delta\hat{U}_{n,\alpha}(w) \leq 0 \tag{3.89}$$

and otherwise use data compressor 2. By (3.80)

$$\Delta U_{0,\alpha}(w) = 0 \tag{3.90}$$

for all $w \in (-\infty,\infty)$, which implies that the optimal policy for the 1 stage problem is a connected policy with threshold

$$T_{1,\alpha} = 0^- \text{ or } \infty \ . \tag{3.91}$$

Now suppose $\Delta U_{n,\alpha}(w)$ is a nondecreasing function of w, for all $w \in (-\infty,\infty)$, and for all $\alpha \in (0,1)$. By (3.88), $\Delta\hat{U}_{n,\alpha}(w)$ is a nondecreasing function of w. Hence, since $t_1 - t_2 + \mu(D_1 - D_2) + \alpha\, \Delta\hat{U}_{n,\alpha}(w)$ is a nondecreasing function of w, equation (3.89) implies that the optimal policy for the first stage of the n+1 stage problem with discount factor $\alpha$ is a connected policy with a threshold labelled $T_{n+1,a} \geq 0^-$. So, from (3.82) and (3.86), $\Delta U_{n+1,\alpha}(w)$ is given by

$$\Delta U_{n+1,\alpha}(w) = \begin{cases} 0 & , \quad \text{if} \quad w \leq -t_1 \\ U_{n+1,\alpha}(w+t_1)-U_{n+1,\alpha}(0) & , \quad \text{if} \quad -t_1 < w \leq -t_2 \\ t_1-t_2+\alpha\Delta\hat{U}_{n,\alpha}(w+t_1) & , \quad \text{if} \quad -t_2 < w \leq T_{n+1,\alpha} - t_1 \\ \mu(D_2 - D_1) & , \quad \text{if} \quad T_{n+1,\alpha}-t_1 < w \leq T_{n+1,\alpha}-t_2 \\ t_1-t_2+\alpha\Delta\hat{U}_{n,\alpha}(w+t_2) & , \quad \text{if} \quad T_{n+1,\alpha} - t_2 < w \end{cases} \tag{3.92}$$

Since $U_{n+1,\alpha}(w)$ and $\Delta\hat{U}_{n,\alpha}(w)$ are nondecreasing functions of w and clearly $\Delta U_{n+1,\alpha}(w)$ is nondecreasing between regions, then $\Delta U_{n+1,\alpha}(w)$ is a nondecreasing function of w. Thus, by induction $\Delta U_{n,\alpha}(w)$ and $\Delta\hat{U}_{n,\alpha}(w)$ are nondecreasing functions of w for all $n \geq 0$ and all $\alpha \in (0,1)$.

Now, from (3.46)

$$\begin{aligned} \Delta V_\alpha(w,\pi^*_\alpha) &= V_\alpha(w + t_1, \pi_\alpha) - V_\alpha(w + t_2, \pi^*_\alpha) \\ &= \lim_{n\to\infty} \Delta U_{n,\alpha}(w) \end{aligned} \tag{3.93}$$

and so $\Delta V_\alpha(w,\pi^*_\alpha)$ is a limit of nondecreasing functions. Thus, $\Delta V_\alpha(w,\pi^*_\alpha)$ is a nondecreasing function of w. Furthermore, from (3.44) and (3.89), $\pi^*_\alpha$ is the policy which selects data compressor 1 if the waiting time w satisfies

$$t_1 - t_2 + \mu(D_1 - D_2) + \alpha \int_0^\infty \Delta V_\alpha(w-\hat{w})F_\tau(d\hat{w}) \leq 0 \qquad (3.94)$$

and otherwise utilizes data compressor 2. But the integral in (3.94) is a decreasing function of w, and so $\pi^*_\alpha$ is a connected policy with a threshold to be labelled $T_\alpha \geq 0^-$.

Q.E.D

Thus, the α-optimal policy has been shown to be a connected policy specified by a threshold $T_\alpha \geq 0^-$. Since all connected policies of this form with negative thresholds are equivalent (policies which always use data compressor 2), then assume $T_\alpha$ is bounded by $T_\alpha \geq -1$. The following lemma proves that $T_\alpha$ is upper bounded for all $\alpha \in (\alpha^*,1)$ where $\alpha^* \in (0,1)$.

Lemma 3.6

The threshold $T_\alpha$ specified in Lemma 3.5 is bounded for all $\alpha \in (\alpha^*,1)$ by

$$T_\alpha < M_T \qquad (3.95)$$

where

$$\alpha^* = \max\left(0, 1 - \frac{(t_1 - t_2)}{2\mu(D_2 - D_1)}\right)$$

and

$$M_T = \max(\lambda^{-1}, (t_1 - t_2) + \frac{\ln 2}{\lambda \ln \alpha^*}) \tag{3.96}$$

Proof

Let $\pi(T)$ be the connected policy with associated threshold T. Let $T_1$ and $T_2$ be such that

$$T_1 + 2(t_1 - t_2) + \mu(D_2 - D_1) < T_2 \tag{3.97}$$

Furthermore, define the sequences of random variables $\{w_n^1\}$ and $\{w_n^2\}$ by

$$w_{n+1}^j = \begin{cases} \max(0, w_n^j + t_1 - \tau_{n+1}) & , \text{ if } w_n \leq T_j \\ \max(0, w_n^j + t_2 - \tau_{n+1}) & , \text{ if } w_n > T_j \end{cases} \quad , j = 1, 2 \tag{3.98}$$

where for j = 1, 2

$$w_0^j = w \quad , \tag{3.99}$$

and $\{\tau_n\}$ is a sequence of independent identically distributed exponential random variables with mean $\lambda^{-1}$. Thus, the sequences defined in (3.98) are the sequences of waiting times under policy $\pi(T_1)$ and $\pi(T_2)$. Then

$$V_\alpha(w,\pi_{T_j}) = \limsup_{N\to\infty} \sum_{n=0}^{N-1} \alpha^n E\{c_j(w_n^j)\} \tag{3.100}$$

where

$$c_j(w) = w + (t_1 + \mu D_1) I(w \leq T_j) + (t_2 + \mu D_2) I(w > T_j) \; . \tag{3.101}$$

To show that $V_\alpha(w,\pi_{T_1})$ is smaller than $V_\alpha(w,\pi_{T_2})$, first it is required to show that $w_n^{(1)} \le w_n^{(2)}$ for all $n \ge 0$. Suppose $w_n^1 \le w_n^2$. Then if $w_n^2 \le T_2$, clearly $w_{n+1}^1 \le w_{n+1}^2$ by (3.98), and if $w_n^2 > T_2$ and $w_n^1 > T_1$, again $w_{n+1}^1 \le w_{n+1}^2$ by (3.98). The case remaining is for $w_n^1 \le T_1$ and $w_n^2 > T_2$. For this case by (3.97) $w_1 + t_1 \le w_2 + t_2$, and so $w_{n+1}^1 \le w_{n+1}^2$. Thus, if $w_n^1 \le w_n^2$ then $w_{n+1}^1 \le w_{n+1}^2$. Therefore, by induction $w_n^1 \le w_n^2$ holds for all n.

Now let

$$N_n = \begin{cases} \max\{0, \sup\{\hat{n}: \hat{n} \le n \text{ and } w_{\hat{n}}^1 = 0\}\} & , \text{ if } n > 0 \\ 0 & , \text{ if } n = 0 \end{cases} \tag{3.102}$$

and

$$A_n = \begin{cases} 1 & \text{if } w_n^1 \ge T_1,\ w_n^2 < T_2 \\ -1 & \text{if } w_n^1 < T_1,\ w_n^2 \ge T_2 \\ 0 & \text{otherwise} \end{cases} . \tag{3.103}$$

Clear if m satisfies $N_n < m \le n$, then $w_m^1 \ne 0$ which implies $w_m^2 \ne 0$. Thus, solving the recurrence relationship (3.98), $w_n^j$ is given by

$$w_n^j = w_{N_n}^j + \sum_{m=N_n}^{n-1} [t_1\, I(w_m^j < T_j) + t_2\, I(w_m^j \ge T_j) - \tau_{m+1}] . \tag{3.104}$$

Then using (3.103) and (3.104), the difference between $w_n^2$ and $w_n^1$ is given by

$$\Delta w_n = w_n^2 - w_n^1$$

$$= \Delta w_{N_n} + \sum_{m=N_n}^{n-1} A_m(t_1 - t_2) . \tag{3.105}$$

It easily follows that if $N_n \neq n$ and $\Delta w_{n-1} \geq T_2 - T_1 - (t_1 - t_2)$, then $\Delta w_n \geq T_2 - T_1 - (t_1 - t_2)$. Hence, by induction, when $\Delta w^2_{N_n} \geq T_2 - T_1 - (t_1 - t_2)$, then $\Delta w_n \geq T_2 - T_1 - (t_1 - t_2)$ and, by (3.97) and (3.101) the difference in costs is bounded by

$$c_2(w^2_n) - c_1(w^1_n) \geq 0 \quad . \tag{3.106}$$

In a similar manner, if $\Delta w_{N_n} < T_2 - T_1 - (t_1 - t_2)$, equation (3.105) implies

$$\sum_{m=N_n}^{n-1} A_m \geq 0 \quad . \tag{3.107}$$

Now define the sequence $B_n$ by

$$B_n = \begin{cases} 0 \quad , & \text{if} \quad \Delta w_{N_n} \geq T_2 - T_1 - (t_1 - t_2) \\ A_n \quad , & \text{otherwise} \ . \end{cases} \tag{3.108}$$

So, from (3.101), (3.105), (3.106), the difference between the cost is bound by

$$\Delta c(w^2_n, w^1_n) = c_2(w^2_n) - c_1(w^1_n)$$

$$\geq \{ \sum_{m=N_n}^{n-1} B_m(t_1 - t_2) + B_n(t_1 - t_2 + \mu(D_1 - D_2)\} \quad . \tag{3.109}$$

Then multiplying (3.109) by $\alpha^n$ and sum over n, the following bound is obtained:

$$\sum_{n=0}^{N-1} \alpha^n \Delta c(w^2_n, w^1_n) \geq \sum_{n=0}^{N-1} \alpha^n \{ \sum_{m=N_n}^{n} B_m(t_1 - t_2) + B_n \mu(D_1 - D_2)\} \tag{3.110}$$

In order to interchange the order of summation in (3.111) define $\hat{N}_n$ by

$$\hat{N}_n = \inf\{\hat{n} > n: \; w^1_{\hat{n}} = 0\} \tag{3.111}$$

and so (3.110) becomes

$$\sum_{n=0}^{N-1} \alpha^n \, \Delta c(w^2_n, w^1_n) \geq \sum_{n=0}^{N-1} B_n \{ \sum_{m=n}^{\min(N-1,\hat{N}_n-1)} \alpha^m (t_1 - t_2) + \alpha^n \mu(D_1 - D_2)\} \, . \tag{3.112}$$

Hence, from (3.100) it is clear that

$$V_\alpha(w,\pi_{T_2}) - V_\alpha(w,\pi_{T_1}) \geq \limsup_{N\to\infty} E\left(\sum_{n=0}^{N-1} \alpha^n B_n \left\{ \left(\frac{1-\alpha^{(\hat{N}_n-n)}}{1-\alpha}\right) (t_1 - t_2) + \mu(D_1 - D_2)\right\}\right) . \tag{3.113}$$

To show that the right side of (3.113) is positive, first define $M_n$ by

$$M_n = \inf\{\hat{n} > n: \; B_{\hat{n}} \neq 0\} \; . \tag{3.114}$$

Then, it is easily shown if n is select such that $B_n = 1$ and $B_{M_n} = -1$, then $M_n < \hat{N}_n$. Hence,

$$\alpha^n B_n\{(\frac{1-\alpha^{\hat{N}_n-n}}{1-\alpha}) (t_1 - t_2) + \mu(D_1 - D_2)\}$$
$$+ \alpha^{M_n} B_{M_n}\{(\frac{1-\alpha^{(\hat{N}_n-M_n)}}{1-\alpha}) (t_1 - t_2) + \mu(D_1 - D_2)\} \geq 0 \, . \tag{3.115}$$

Furthermore, it is clear that if n is selected such that $B_n = -1$, then $B_{M_n} = 1$. Thus, for every $\hat{n}$ such that $B_{\hat{n}} = -1$ there exists a $n < \hat{n}$ such that $B_n = 1$, $M_n = \hat{n}$, and (3.115) is satisfied. Now define the sequence $C_n$ by

$$C_n = \begin{cases} 0 & \text{if } B_n = 1, B_{M_n} = -1 \\ 0 & \text{if } B_n = -1 \\ B_n & \text{otherwise} \end{cases} \quad . \qquad (3.116)$$

Thus, using (3.115), (3.113) can be further lower bounded by

$$V_\alpha(w,\pi_{T_2}) - V_\alpha(w,\pi_{T_1}) \geq \limsup_{N\to\infty} E \sum_{n=0}^{N-1} \alpha^n C_n\{(\frac{1-\alpha^{\hat{N}_n-n}}{1-\alpha})(t_1 - t_2)$$
$$+ \mu(D_1 - D_2)\} \quad . \qquad (3.117)$$

Clearly from (3.116), $C_n = 0$ or 1. Thus, to prove the right side of (3.117) is positive, it is sufficient to prove that for all $n \geq 0$

$$E\{(\frac{1-\alpha^{(\hat{N}_n-n)}}{1-\alpha})(t_1 - t_2) + \mu(D_1 - D_2)|C_n = 1\} \geq 0 \quad . \qquad (3.118)$$

Using Jensen's inequality, a lower bound to the right side of (3.118) is given by

$$E\{(\frac{1-\alpha^{(\hat{N}_n-n)}}{1-\alpha})(t_1 - t_2) + \mu(D_1 - D_2)|C_n = 1\}$$
$$\geq (\frac{1-\alpha^{E(\hat{N}_n-n|C_n=1)}}{1-\alpha})(t_1 - t_2) + \mu(D_1 - D_2) \quad . \qquad (3.119)$$

Now $E(\hat{N}_n - n|C_n = 1)$ can be lower bounded by examining the sequence

$$\tilde{w}_{n+1} = \tilde{w}_n - \tau_{n+1} \qquad (3.120)$$

where $w_0 = T_1 - (t_1 - t_2)$ and $\{\tau_n\}$ is a sequence of independent identical distributed exponential random variables with mean $\lambda^{-1}$. Hence, if

$$\tilde{N} = \inf\{n \geq 0: \tilde{W}_{n+1} \leq 0\} \quad , \tag{3.121}$$

then it is easily argued that

$$E(\tilde{N}) \leq E(\hat{N}_n - n | C_n = 1) \tag{3.122}$$

But $\tilde{N}$ is a Poisson random variable with mean $\lambda(T_1 - (t_1 - t_2))$.

Hence, using (3.122) in (3.119), it is found that

$$E\{(\frac{1-\alpha^{(\hat{N}_n - n)}}{1-\alpha})(t_1 - t_2) + \mu(D_1 - D_2)|C_n = 1\}$$

$$\geq (\frac{1-\alpha^{\lambda(T_1-(t_1-t_2))}}{1-\alpha})(t_1 - t_2) + \mu(D_1 - D_2) \tag{3.123}$$

Now let

$$\alpha^* = \max\left(0, 1 - \frac{(t_1 - t_2)}{2\mu(D_2 - D_1)}\right) \tag{3.124}$$

and

$$T_1 = \max(\lambda^{-1}, (t_1 - t_2) + \frac{\ln 2}{\lambda \ln \alpha^*}) < \infty \quad . \tag{3.125}$$

So, it is easily shown for $\alpha \in (\alpha^*, 1)$,

$$(\frac{1-\alpha^{\lambda(T_1-(t_1-t_2))}}{1-\alpha})(t_1 - t_2) + \mu(D_1 - D_2) \geq 0 \quad . \tag{3.126}$$

Hence, from (3.126) and (3.123), equation (3.118) is valid for $\alpha \in (\alpha^*, 1)$ when $\alpha^*$ is given by (3.124) and $T_1$ is given by (3.125). Then substituting (3.118) into (3.117) it is clear that for $\alpha \in (\alpha^*, 1)$

$$V_\alpha(w, \pi_{T_2}) - V_\alpha(w, \pi_{T_1}) \geq 0 \quad . \tag{3.127}$$

AD-A035 222 CALIFORNIA UNIV LOS ANGELES SCHOOL OF ENGINEERING A--ETC F/G 17/2
SOURCE CODING AND ADAPTIVE DATA COMPRESSION FOR COMMUNICATION N--ETC(U)
DEC 76 H S NUSSBAUM N00014-75-C-0609

UNCLASSIFIED UCLA-ENG-76121 NL

2 OF 3

AD
A035222

Thus, $\pi_{T_2}$ can not be α-optimal, and so, from (3.97) the α-optimal threshold $T_\alpha$ must satisfy

$$T_\alpha \le T_1 + 2(t_1 - t_2) + \mu(D_2 - D_1) < \infty \quad . \tag{3.128}$$

<u>Q.E.D</u>

Before proceeding with the verification of Assumptions 3.2 and 3.3, consider the sequences of random variables generated by the following equations: for $n \geq 0$

$$\hat{\hat{w}}_{n+1} = \max(0, \hat{\hat{w}}_n + t_2 - \tau_{n+1}) \tag{3.129}$$

and

$$\hat{w}_{n+1} = \begin{cases} 0 & , \text{if} \quad \hat{\hat{w}}_n = 0 \text{ and } \tau_{n+1} > z + t_2 \\ \hat{\hat{w}}_{n+1} + z & , \text{otherwise} \end{cases} \tag{3.130}$$

where $\hat{\hat{w}}_0 = \hat{\hat{w}} \geq 0$, $\hat{w}_0 = \hat{\hat{w}}_0 + z$, $\{\tau_n\}$ is a sequence of independent identically distributed exponential random variables with mean $\lambda^{-1}$, z is an arbitrary number, and $t_2$ is given by (3.22). The sequence of random variables $\{\hat{\hat{w}}_n\}$ can be considered a sequence of waiting times of a M/D/1 queueing system with service time $t_2$ and mean interarrival time equal to $\lambda^{-1}$. Thus, the sequence of random variables $\{\hat{w}_n\}$ is observed to be the shifted version of $\{\hat{\hat{w}}_n\}$, except that after two consecutive values of $\hat{\hat{w}}_n$ are zero, $\hat{w}_n$ may transition to a state labelled 0.

Now consider the sequences of waiting times in two separate M/D/1 queueing systems which use connected policies with thresholds $T_1$ and

$T_2$. Let $\{w_n^{(0,1)}\}$ and $\{w_n^{(1,1)}\}$ be the sequences of waiting times when threshold $T_1$ is used where $w_0^{(0,1)} = 0$ and $w_0^{(1,1)} = w$. Furthermore, let $\{w_n^{(0,2)}\}$ and $\{w_n^{(1,2)})\}$ be the waiting time sequences when threshold $T_2$ is used where $w_0^{(0,2)} = 0$ and $w_0^{(1,2)} = w$. So, from (3.21) the recurrence relationship for the waiting time sequences are given by

$$w_{n+1}^{(i,j)} = \begin{cases} \max(0,\ w_n^{(i,j)} + t_1 - \tau_{n+1}) \ , \text{ if } & w_n^{(i,j)} \le T_j \\ \max(0,\ w_n^{(i,j)} + t_2 - \tau_{n+1}) \ , \text{ if } & w_n^{(i,j)} > T_j \end{cases} \tag{3.131}$$

where

$$w_0^{(0,j)} = 0 \quad , \quad j = 1, 2 \quad , \tag{3.132}$$

and

$$w_0^{(1,j)} = w \ge 0 \ , \quad j = 1, 2 \quad , \tag{3.133}$$

and $\{\tau_n\}$ is the same sequence as in (3.129). The following lemma shows $\hat{w}_n$ of (3.130) upper bounds $w_n^{(i,j)}$.

<u>Lemma 3.7</u>

For the sequences defined by (3.130) and (3.131), for all $n \ge 0$

$$w_n^{(i,j)} \le \hat{w}_n \quad , \quad i = 0, 1, \quad j = 1, 2 \tag{3.134}$$

where

$$z = T_1 + t_1 \ , \tag{3.135}$$

$$w_0^{(0,1)} = w_0^{(0,2)} = 0 \ , \tag{3.136}$$

$$w_0^{(1,1)} = w_0^{(1,2)} = w \ , \tag{3.137}$$

and

$$\hat{\hat{w}}_0 = \max(0, w-z) \quad . \tag{3.138}$$

<u>Proof</u>

The proof is by induction. First, fix i and j. From (3.136), (3.137) and (3.138) it is clear that

$$w_0^{(i,j)} \le \hat{w}_0 \quad . \tag{3.139}$$

Now suppose $w_n^{(i,j)} \le \hat{w}_n$. Then for $\hat{w}_n \ge z$,

$$\hat{w}_n + t_2 - \tau_{n+1} \ge \begin{cases} w_n^{(i,j)} + t_1 - \tau_{n+1} \ , & \text{if} \quad w_n^{(i,j)} \le T_j \\ w_n^{(i,j)} + t_2 - \tau_{n+1} \ , & \text{if} \quad w_n^{(i,j)} > T_j \end{cases} \tag{3.140}$$

since $w_n^{(i,j)} + t_1 < z$ when $w_n^{(i,j)} \le T_j$. Then, using (3.131),

$$\max(\hat{\hat{w}}_n + t_2 - \tau_{n+1}, 0) + z \ge w_{n+1}^{(i,j)} \tag{3.141}$$

and

$$\max(\hat{w}_n + t_2 - \tau_{n+1}, 0) \ge w_{n+1}^{(i,j)} \quad . \tag{3.142}$$

Thus, from (3.129) and (3.130)

$$\hat{w}_{n+1} \ge w_{n+1}^{(i,j)} \quad .$$

For the remaining case of $\hat{w}_n = 0$, equation (3.139) gives $w_n^{(i,j)} = 0$. Then from (3.129) and (3.130)

$$w_{n+1}^{(i,j)} \le \begin{cases} 0\ , & \text{if} \quad \hat{w}_{n+1} = 0 \\ z\ , & \text{if} \quad \hat{w}_{n+1} \ge z \end{cases} \tag{3.143}$$

and so,

$$\hat{w}_{n+1} \ge w_{n+1}^{(i,j)} \quad . \tag{3.144}$$

Consequently, by induction the result (3.134) is achieved.

Q.E.D

Using random variables $\{\hat{w}_n\}$ the average waiting time of the queueing systems can be bounded. In order to bound the performance of the system, it is required to find the mean number of steps $\hat{T}_z(w)$ for the sequence of random variables $\{\hat{w}_n\}$ to reach state 0 when $\hat{\hat{w}}_0 = \max(0, w-z)$.

**Lemma 3.8**

For $w \ge z$, $\hat{T}_z(w)$ satisfies

$$\hat{T}_z(w) = 1 + [e^{\lambda z} + \lambda(w-z)]\ (1 - \lambda t_2)^{-1} \tag{3.145}$$

**Proof**

Define the sequence $\{S_n\}$ by

$$S_n = \sum_{i=1}^{n} (\tau_i - t_2) \qquad n \ge 1 \tag{3.146}$$

where

$$S_0 = 0 \tag{3.147}$$

and $\{\tau_n\}$ is the same sequence as in (3.129). Let N be the stopping time given by

$$N = \inf\{n: \; S_n \geq w-z\} \tag{3.148}$$

which is the number of steps to get from $\hat{w}_0 = w$ to state z. Wald's equation gives

$$E\{\sum_{i=1}^{N} (\tau_i - t_2)\} = E(N)\; E(\tau_1 - t_2) \quad . \tag{3.149}$$

But by appropriately adding and subtracting w-z, it is obtained that

$$E\{\sum_{n=1}^{N} (\tau_n - t_2)\} = w-z + E\left\{E\left(\tau_N - [w-z-\sum_{n=1}^{N-1} (\tau_n - t_2) + t_2] \middle| N,\{\tau_n,\; n = 1,2,\ldots,N-1\}\right)\right\} \tag{3.150}$$

Using the memoryless properties of an exponential random variable, the inner expectation of (3.150) becomes

$$E\left\{\tau_N - [w - z - \sum_{n=1}^{N-1} (\tau_n - t_2) + t_2] \middle| N,\{\tau_n,\; n = 1,2,\ldots,N-1\}\right\} = \lambda^{-1}. \tag{3.151}$$

Hence, equation (3.150) becomes

$$E\{\sum_{n=1}^{N} (\tau_n - t_2)\} = w - z + \lambda^{-1} \quad . \tag{3.152}$$

Thus, using (3.149), the mean number of steps $\hat{T}_1(w)$ to go from $\hat{w}_0 = w$ to z is given by

$$\hat{T}_1(w) = E(N)$$

$$= \frac{w - z + \lambda^{-1}}{\lambda^{-1} + t_2} \quad . \tag{3.153}$$

Now let p be defined as

$$p = \Pr\{\hat{w}_{n+1} = 0 | \hat{w}_n = z)$$

$$= e^{-\lambda z} \quad . \tag{3.154}$$

Furthermore, the mean number of customers served $N_B$ in a busy period of the M/D/1 queueing system generating the sequence $\{\hat{\hat{w}}_n\}$ is equal to (Cohen [36])

$$E(N_B) = (1 - t_2\lambda)^{-1} \quad . \tag{3.155}$$

So, the mean number of steps, $T_2$, to go from $\hat{w}_n = z$ to state 0 is given by

$$\hat{T}_2 = 1 + p \sum_{n=0}^{\infty} (1-p)^n \, n \, E(N_B)$$

$$= 1 + (e^{\lambda z} - 1)(1 - t_2\lambda)^{-1} \quad . \tag{3.156}$$

Finally, since $\hat{T}_z(w)$ is the sum of $\hat{T}_1(w)$ and $\hat{T}_2$, $\hat{T}_z(w)$ is given by

$$\hat{T}_z(w) = \hat{T}_1(w) + \hat{T}_2$$

$$= 1 + (e^{\lambda z} + \lambda(w-z))(1 - t_2\lambda)^{-1} \quad . \tag{3.157}$$

<u>Q.E.D</u>

Assumption 3.2 can now be verified.

Lemma 3.9

For $\alpha^* = \max\left(0,\ 1 - \frac{t_1 - t_2}{2\mu(D_2 - D_1)}\right)$ and for all $\alpha \in (\alpha^*, 1)$,

$$|V_\alpha(w, \pi_\alpha^*) - V_\alpha(w, \pi_\alpha^*)| \le L(w) \tag{3.158}$$

where

$$L(w) = \{1 + (e^{\lambda(M_T + t_1)} + \lambda w)\}\,(1 - t_2\lambda)$$
$$\cdot \left\{w + 2(t_1 - t_2) + \mu(D_2 - D_1)\right\} \tag{3.159}$$

and $M_T$ is given by Lemma 3.6.

Proof

Let $\alpha^* = \max\left(0,\ 1 - \frac{t_1 - t_2}{2\mu(D_2 - D_1)}\right)$ and $\alpha \in (\alpha^*, 1)$. Consider the sequences $\{w_n^{(i,j)}\}$ defined in (3.131) where the parameters $T_1$ and $T_2$ are given by $T_1 = M_T$, $T_2 = T_\alpha$ and $M_T$ is given in Lemma (3.6). So, Lemma 3.7 implies that if $\hat{w}_n = 0$ then for $i = 0, 1$ and $j = 1, 2$, $w_n^{(i,j)} = 0$. Furthermore, let

$$\hat{N} = \inf\{n \ge 0 :\ \hat{w}_n = 0\} \tag{3.160}$$

and so, from (3.131)

$$w_n^{(1,j)} - w_n^{(0,j)} = 0 \qquad \text{for} \qquad n = \hat{N}, \hat{N}+1, \ldots \tag{3.161}$$

Since $T_2 = T_\alpha$ and $T_\alpha$ is the threshold generating $\pi_\alpha^*$, then (3.100) and

the definition of $c_j(w)$ in (3.101) implies

$$V_\alpha(w,\pi_\alpha^*) - V_\alpha(w,\pi_\alpha^*) = E \sum_{n=0}^{\hat{N}} \alpha^n \{c_2(w_n^{(1,2)}) - c_2(w_n^{(0,2)})\} \quad . \tag{3.162}$$

Now it is easily observed that

$$|w_n^{(1,2)} - w_n^{(0,2)}| \le w + t_1 - t_2 \tag{3.163}$$

and so

$$|c_2(w_n^{(1,2)}) - c_2(w_n^{(0,2)})| \le w + 2(t_1 - t_2) + \mu(D_2 - D_1) \quad . \tag{3.164}$$

Thus, (3.164) is bounded by

$$\begin{aligned} |V_\alpha(w,\pi_\alpha^*) - V_\alpha(w,\pi_\alpha^*)| &\le E\Big( \sum_{n=0}^{\hat{N}} \alpha^n |c_2(w_n^{(1,2)}) - c_2(w_n^{(0,2)})| \\ &\le E(\hat{N})\ (w + 2(t_1 - t_2) + \mu(D_2 - D_1)) \\ &\le \hat{T}_{M_T+t_1}(\max(w, M_T + t_1)) \\ &\quad \cdot \{w + 2(t_1 - t_2) + \mu(D_2 - D_1)\} \quad . \end{aligned} \tag{3.165}$$

Using Lemma 3.8 for the expression for $\hat{T}_z(w)$,

$$\begin{aligned} |V_\alpha(w,\pi_\alpha^*) - V_\alpha(0,\pi_\alpha^*)| &\le \{1 + [e^{\lambda(M_T+t_1)} + \lambda(\max(w - M_T - t_1, 0))\} \\ &\quad \cdot (1 - \lambda t_2)^{-1} \{w + 2(t_1 - t_2) + \mu(D_2 - D_1)\} \end{aligned} \tag{3.166}$$

and (3.158) is achieved when it is noted that

$$\max(w - M_T - t_1, 0) \le w \quad . \tag{3.167}$$

<u>**Q.E.D**</u>

To verify the final assumption, it is required to determine if a connected policy with threshold T induces a limiting probability distribution of the waiting time, and if the limit of the mean waiting time exists and is finite. Lindley's results (Lindley [33]) pertaining to the properties of the limit of the waiting time distribution apply to queueing system for which each service is statistically independent of prior service times and prior arrival times. Thus, these results are not applicable to the present situation and the following lemma provides the necessary results.

**Lemma 3.10**

For $\lambda t_2 < 1$, the distribution function of the waiting time $w_n$, under a connected policy with threshold $T < \infty$, has a limit given by

$$F(W) = \lim_{n\to\infty} \Pr(w_n \leq W \mid w_0 = w) \tag{3.168}$$

where F(W) is independent of w. Furthermore F(W) is a proper distribution, and $\lim_{n\to\infty} E(w_n)$ exists and is finite.

**Proof**

Consider equations (3.129) and (3.130). Since $\hat{\hat{w}}_n$ represents the waiting time in a M/D/1 queueing system with traffic intensity less than 1 (i.e., $\lambda t_2 < 1$), Lindley's theorem (Lindley [33]) indicates that $\Pr(\hat{w}_n = 0)$ approaches a limit as $n \to \infty$. But

$$\Pr(\hat{w}_{n+1} = 0) = \Pr(\hat{\hat{w}}_n = 0)\ \Pr(\tau_{n+1} \geq z + t_2) \tag{3.169}$$

and so, $\Pr(\hat{w}_n = 0)$ approaches a limit, as $n \to \infty$, which is denoted by

$$\hat{P}_0 = \lim \Pr(\hat{w}_n = 0) \quad . \tag{3.170}$$

Let the sequences $\{\hat{w}_n\}$ and $\{w_n^{(i,j)}\}$ be defined as in Lemma 3.7 with $T_1 = T$ and $T_2 = 0$. Then $\Pr(w_n^{(i,1)} \leq W)$, $W \geq 0$, is obtained by conditioning on the events $\hat{w}_\ell = 0$, $\ell = 0, 1, \ldots, n$, and so,

$$\Pr(w_n^{(i,1)} \leq W) = \sum_{k=0}^{n} \Pr(w_n^{(i,1)} \leq W;\ \hat{w}_\ell > 0,\ \ell = k+1,\ldots,n-1 \mid \hat{w}_k = 0)$$
$$\cdot \Pr(\hat{w}_k = 0) + \Pr(w_n^{(i,1)} \leq W;\ \hat{w}_\ell > 0,\ \ell = 0,1,\ldots,n) \tag{3.171}$$

To show that the limit as $n \to \infty$ of $\Pr(w_n^{(i,1)} \leq W)$ exists, first consider the quantity $P_n$ given by

$$P_n = \Pr(\hat{w}_\ell > 0,\ \ell = 1,2,\ldots,n-1 \mid \hat{w}_0 = 0)$$
$$= \sum_{k=n}^{\infty} \Pr(\hat{w}_k = 0;\ \hat{w}_\ell > 0,\ \ell = 1,2,\ldots,k-1 \mid \hat{w}_0 = 0) \ , \tag{3.172}$$

for $n > 1$ and $P_1 = P_0 = 1$. Hence, it is clear that

$$\Pr(w_n^{(i,j)} \leq W;\ \hat{w}_\ell > 0,\ \ell = k+1,\ldots,n-1 \mid \hat{w}_k = 0) \leq P_{n-k} \tag{3.173}$$

Now summing $P_n$ over n gives

$$\sum_{n=1}^{N} P_n = N \sum_{k=N+1}^{\infty} \Pr(\hat{w}_k = 0,\ \hat{w}_\ell > 0,\ \ell = 1,2,\ldots,i-1 \mid \hat{w}_0 = 0)$$
$$+ \sum_{n=1}^{N} n \Pr(\hat{w}_n = 0,\ \hat{w}_\ell > 0,\ \ell = 1,2,\ldots,n-1 \mid \hat{w}_0 = 0) \ . \tag{3.174}$$

But the mean number of steps $\hat{T}$ to go from $\hat{w}_0 = 0$ back to state 0 is given by

$$\hat{T} = \lim_{N\to\infty} \sum_{n=1}^{N} n \Pr(\hat{w}_n = 0;\ \hat{w}_\ell > 0,\ \ell = 1,2,\ldots,n-1 | \hat{w}_0 = 0) \quad (3.175)$$

and $\hat{T}$ is finite from Lemma 3.8. So, using (3.174) and (3.175),

$$\hat{T} - \sum_{n=1}^{N} P_n = \sum_{n=N+1}^{\infty} (n-N) \Pr(\hat{w}_n = 0,\ \hat{w}_\ell > 0,\ \ell = 1,2,\ldots,i-1 | \hat{w}_0 = 0) \quad (3.176)$$

Taking the limit as $N \to \infty$ of (3.176) and noting that the summation in (3.175) has a limit, $\hat{T}$ is given by

$$\hat{T} = \lim_{N\to\infty} \sum_{n=1}^{N} P_n \ . \quad (3.177)$$

Thus, for every $\varepsilon > 0$ there exists a $N(\varepsilon)$ such that

$$\sum_{k=n}^{N} P_k < \varepsilon \quad (3.178)$$

for $N > n \geq N(\varepsilon)$, and clearly, from (3.170) for every $\varepsilon > 0$ there exists a $M(\varepsilon)$ such that

$$|\Pr(\hat{w}_n = 0) - \hat{P}_0| < \varepsilon \quad (3.179)$$

for $n \geq M(\varepsilon)$.

Now define $F_N(W)$ by

$$F_N(W) = \hat{P}_0 \sum_{k=0}^{N} \Pr(w_k^{(0,1)} \leq W;\ \hat{w}_\ell > 0,\ \ell = 1,2,\ldots,k-1 | w_0^{(0,1)} = 0). \quad (3.180)$$

Then from (3.171) and (3.173) the difference between $\Pr(w_{n+N}^{(i,1)} \leq W)$ and $F_N(W)$ is bounded by

$$|Pr(w^{(i,1)}_{n+N} \le W) - F_N(w)| \le \sum_{k=0}^{N} |Pr(\hat{w}_{N+n-k} = 0) - \hat{P}_0| P_k$$

$$+ \sum_{k=N+1}^{N+n} Pr(\hat{w}_{N+n-k} = 0)\, P_k + Pr(w^{(i,1)}_n \le W;\ \hat{w}_\ell > 0, \ell = 0,1,\ldots,n+N) \tag{3.181}$$

From Lemma 3.8 it is easily established that for all $\varepsilon > 0$ there exists a function $K(\varepsilon)$ such that for all $n > K(\varepsilon)$ and $i = 0,1$

$$Pr(w^{(i,1)}_n \le W;\ \hat{w}_\ell > 0,\ \ell = 0,1,\ldots,n+N) < \varepsilon \quad . \tag{3.182}$$

Now let $N > \max(N(\varepsilon/3), M(\varepsilon/3\hat{T}), K(\varepsilon/3))$, then using (3.178), (3.179) and (3.182) in (3.181), it is clear that for $n > 0$ and $i = 0,1$

$$|Pr(w^{(i,1)}_{n+N} < W) - F_N(W)| \le \sum_{k=0}^{N} \varepsilon P_k/3\hat{T} + \varepsilon/3 + \varepsilon/3$$

$$\le \varepsilon \tag{3.183}$$

where the last inequality is derived from (3.177). Thus, the limit of the distribution $Pr(w^{(i,1)}_n < W)$ exists, is independent of i which implies it is independent of $w^{(i,1)}_0$, and is denoted by

$$F(W) = \lim_{n\to\infty} Pr(w^{(i,1)}_n \le W) \tag{3.184}$$

To show that F(W) is a valid distribution function, consider $\lim_{n\to\infty} Pr(\hat{w}_n \le W)$. It is easily shown, using Lindley's Theorem, that this limit exists and is a valid distribution function which is denoted

$$\hat{F}(W) = \lim_{n\to\infty} Pr(\hat{w}_n \le W) \tag{3.185}$$

But from Lemma 3.7

$$\hat{w}_n \geq w_n^{(i,j)} \tag{3.186}$$

for all n, and so

$$F(W) \geq \hat{F}(W) \quad . \tag{3.187}$$

Since $\lim_{N\to\infty} \hat{F}(W) = 1$, by (3.187) $\lim_{W\to\infty} F(W) = 1$. Hence, $F(W)$ is a valid distribution function.

The limit of the mean of $w_n^{(i,1)}$ as $n \to \infty$ is shown to exist by noting that

$$E(w_n^{(i,1)}) = \int_0^\infty (1 - \Pr(w_n^{(i,1)} \leq W))\, dW \tag{3.188}$$

and

$$E(\hat{w}_n) = \int_0^\infty (1 - \Pr(\hat{w}_n \leq W))\, dW \tag{3.189}$$

From (3.186) and (3.187) for $W \varepsilon [0,\infty)$

$$1 - \Pr(\hat{w}_n \leq W) \geq 1 - \Pr(w_n^{(i,1)} \leq W) \geq 0 \tag{3.190}$$

and hence

$$1 - \hat{F}(W) \geq 1 - F(W) > 0 \quad . \tag{3.191}$$

But clearly, $E(\hat{w}_n) < \infty$ for all n, and

$$\lim_{n\to\infty} E(\hat{w}_n) = \int_0^\infty (1 - \hat{F}(W))\, dW < \infty \quad . \tag{3.192}$$

Hence, using a version of the dominated convergence theorem (Royden [37], p. 89)

$$\lim_{n\to\infty} E(w_n^{(i,1)}) = \int_0^\infty (1 - F(W))\, dW < \infty \quad . \tag{3.193}$$

Q.E.D

Finally Assumption 3.3 can be verified.

**Lemma 3.11**

There exists an increasing sequence $\{\alpha_n\}$, $\alpha_n \in (0,1)$, such that $\alpha_n \uparrow 1$ and $\lim_{n\to\infty} T_{\alpha_n}$ exists. Furthermore, the connected policy $\hat{\pi}$ with threshold $\hat{T}$ given by

$$\hat{T} = \lim_{n\to\infty} T_{\alpha_n} \tag{3.194}$$

satisfies

$$\Phi(w,\hat{\pi}) = \limsup_{n\to\infty} (1 - \alpha_n)\, V_{\alpha_n}(w,\pi_\alpha^*) \quad , \tag{3.195}$$

for all $w \in [0,\infty)$.

**Proof**

From Lemma 3.5 and Lemma 3.6, there exists a $\alpha^* \in (0,1)$ and a $M_T$ such that

$$-1 \leq T_\alpha < M_T < \infty \qquad \text{for all } \alpha \in (\alpha^*,1) \quad . \tag{3.196}$$

Since $T_\alpha$ is bounded for all $\alpha \in (\alpha^*,1)$, there exists an increasing sequence $\{\alpha_n\}$, $\alpha_n \in (0,1)$, such that $\alpha_n \uparrow 1$ and $\lim_{n\to\infty} T_{\alpha_n}$ exists. Let $\hat{\pi}$ be the connected policy with threshold $\hat{T}$ given by (3.194), for such a sequence $\{\alpha_n\}$.

Now, for all $\alpha \in (0,1)$

$$|\Phi(w,\hat{\pi}) - (1-\alpha)\, V_\alpha(w,\pi_\alpha^*)| \le$$

$$|\Phi(w,\hat{\pi}) - (1-\alpha)\, V_\alpha(w,\hat{\pi})| + (1-\alpha)\, |V_\alpha(w,\hat{\pi}) - V_\alpha(w,\pi_\alpha^*)| \quad . \tag{3.197}$$

Using Lemma 3.10 and (3.39)

$$\Phi(w,\hat{\pi}) = \lim_{n\to\infty} E_{\hat{\pi}}\{c(w_n,a_n)|w_0 = w\}$$

$$= \lim_{n\to\infty} \{E_{\hat{\pi}}(w_n) + t_2 + \mu D_2 + (t_1 - t_2 + \mu(D_1 - D_2)\, \Pr(w_n \le \hat{T})\} \, . \tag{3.198}$$

From (3.37) for all $N' > 0$

$$|V_\alpha(w,\hat{\pi}) - \limsup_{N\to\infty} \sum_{n=N'}^{N} \alpha^n\, \Phi(w,\hat{\pi})| \le \sum_{n=0}^{N'-1} \alpha^n\, E_{\hat{\pi}}(c(w_n,a_n)|w_0 = w)$$

$$+ \limsup_{N\to\infty} \sum_{n=N'}^{N} \alpha^n |E_{\hat{\pi}}(c(w_n,a_n)|w_0 = w) - \Phi(w,\hat{\pi})| \tag{3.199}$$

But from (3.198), for every $\varepsilon > 0$ there exists a $N(\varepsilon)$ such that for all $n \ge N(\varepsilon)$

$$|E_{\hat{\pi}}\{c(w_n,a_n)|w_0 = w\} - \Phi(w,\hat{\pi})| < \varepsilon \quad . \tag{3.200}$$

Also, Lemma 3.3 implies that for all $N \ge 0$, the sequence $B_N$ defined by

$$B_N = \max_{0\le n\le N} E_{\hat{\pi}}(c(w_n,a_n)|w_0 = w) \tag{3.201}$$

is bounded. Thus, using (3.200) and (3.201) in (3.199), the following bound is found when $N' > N(\varepsilon)$:

$$\left|V_\alpha(w,\hat{\pi}) - \frac{\alpha^{N'}}{1-\alpha}\,\Phi(w,\hat{\pi})\right| \le \frac{1-\alpha^{N'}}{1-\alpha}\,B_{N'} + \frac{\alpha^{N'}}{1-\alpha}\,\varepsilon \quad . \tag{3.202}$$

Hence, from (3.202),

$$\left|(1-\alpha)\,V_\alpha(w,\hat{\pi}) - \Phi(w,\hat{\pi})\right| \le (1-\alpha^{N'})\,(\Phi(w,\hat{\pi}) + B_{N'}) + \alpha^{N'}\,\varepsilon \; . \tag{3.203}$$

Now consider the sequences of random variables $\{\hat{w}_n\}$ and $\{w_n^{(1,j)}\}$ which are discussed in Lemma 3.7, where $T_1 = \max(\hat{T},T_\alpha)$, $T_2 = \min(\hat{T},T_\alpha)$, and $\alpha \in (\alpha^*,1)$. So, from (3.100)

$$V_\alpha(w,\pi_{T_j}) = \limsup_{N\to\infty} \sum_{n=0}^{N} \alpha^n\, E\{c_j(w_n^{(1,j)})\} \tag{3.204}$$

where $c_j(w)$ is given by (3.101). So, the difference between the $\alpha$-discounted cost under $\pi_\alpha^*$ and $\hat{\pi}$ is bounded by

$$\left|V_\alpha(w,\pi_\alpha^*) - V_\alpha(w,\hat{\pi})\right| \le \limsup_{N\to\infty} \sum_{n=0}^{N} \alpha^n\, E\{\left|c_1(w_n^{(1,1)}) - c_2(w_n^{(1,2)})\right|\} \; . \tag{3.205}$$

Now define

$$N_0 = \inf\{n:\; w_n^{(1,1)} = w_n^{(1,2)} \text{ and } T_2 < w_n^{(1,1)} \le T_1\}\;, \tag{3.206}$$

$$n_i = \inf\{n > N_i:\; \hat{w}_n = 0\} \qquad i = 0,1,2,\ldots \tag{3.207}$$

and

$$N_i = \inf\{n > n_{i+1}:\; w_n^{(i,1)} = w_n^{(1,2)} \text{ and } T_2 < w_n^{(1,1)} \le T_1\}\;. \qquad i = 1,2,\ldots \tag{3.208}$$

For a fixed value of i, if n satisfies $N_i \le n < n_i$, then the difference

in waiting times is bounded by

$$|w_n^{(1,1)} - w_n^{(1,2)}| \le T_1 - T_2 + t_1 - t_2 \tag{3.209}$$

and so

$$|c_1(w_n^{(1,1)}) - c_2(w_n^{(1,2)})| \le T_1 - T_2 + 2(t_1 - t_2) + \mu(D_2 - D_1). \tag{3.210}$$

In a similar manner, for a fixed value of i, if n satisfies $n_i \le n < N_{i+1}$, then the difference in costs is

$$|c_1(w_n^{(1,1)}) - c_2(w_n^{(1,2)})| = 0 \ . \tag{3.211}$$

Thus, (3.205) can be written

$$\begin{aligned} |V_\alpha(w,\pi_\alpha^*) - V_\alpha(w,\hat{\pi})| &\le \limsup_{N\to\infty} \sum_{i=0}^{N} E\left\{\sum_{n=N_i}^{n_i-1} \alpha^n |c_1(w_n^{(1,1)}) - c_2(w_n^{(1,2)})|\right\} \\ &\le \limsup_{N\to\infty} \sum_{i=0}^{N} E\left\{\sum_{n=N_i}^{n_i-1} \alpha^n \,[T_1-T_2 + 2(t_1-t_2) \right. \\ &\qquad\qquad \left. + \mu(D_2 - D_1)]\right\} \end{aligned} \tag{3.212}$$

Now clearly,

$$E\{\sum_{n=N_i}^{n_i-1} \alpha^n | N_i\} \le \alpha^{N_i} \ E(n_i - 1 - N_i|N_i) \quad i = 0,1,2,\ldots \tag{3.213}$$

and $E(n_i - 1 - N_i|N_i)$ is equivalent to the mean number of steps to go from $w_0^{(i,j)} = w$, where $T_2 < w \le T_1$, to $\hat{w}_n = 0$. So, using Lemma 3.8 and (3.213),

$$E\{\sum_{n=N_i}^{n_i-1} \alpha^n | N_i\} \le \alpha^{N_i} \hat{T}_{T_1+t_1}(T_1 + t_1)$$

$$\le \alpha^{N_i} \hat{T}_{M_T+t_1}(T_1 + t_1) < \infty \qquad (3.214)$$

where $M_T$ is given by Lemma 3.6. Hence, the bound in (3.212) becomes

$$|V_\alpha(w,\pi^*_\alpha) - V_\alpha(w,\hat{\pi})| \le \limsup_{N\to\infty} \sum_{i=0}^{N} E\{\alpha^{N_i}\}B \qquad (3.215)$$

where

$$B = \hat{T}_{M_T+t_1}(M_T + t_1)\ [M_T + 2(t_1 - t_2) + \mu(D_2 - D_1)] < \infty \qquad (3.216)$$

and B is independent of $\alpha$, $T_1$, and $T_2$. Now

$$\limsup_{N\to\infty} \sum_{i=0}^{N} E(\alpha^{N_i}) = \limsup_{N\to\infty} \sum_{n=0}^{N} \alpha^n (\sum_{i=0}^{n} Pr(N_i = n)) \qquad (3.217)$$

and from the definition of $N_i$ in (3.210),

$$\sum_{i=0}^{n} Pr(N_i = n) \le Pr(T_2 < w_n^{(1,1)} \le T_1) \quad . \qquad (3.218)$$

Since the interarrival times are exponential random variables, then it is easily verified that

$$Pr(T_2 < w_n^{(1,1)} \le T_1 | w_{n-1}^{(1,1)}) \le 1 - e^{-\lambda(T_1-T_2)} \qquad (3.219)$$

for an arbitrary value of $w_{n-1}^{(1,1)}$. So, (3.217) is upper bounded by

$$\limsup_{N\to\infty} \sum_{i=0}^{N} E(\alpha^{N_i}) \le \limsup_{N\to\infty} \sum_{n=0}^{N} \alpha^n (1 - e^{-\lambda(T_1-T_2)})$$

$$= \frac{1 - e^{-\lambda(T_1-T_2)}}{1 - \alpha} \quad . \qquad (3.220)$$

Hence, upon substituting (3.220) in (3.215) and noting that

$T_1 = \max(T_\alpha, \hat{T})$ and $T_2 = \max(T_\alpha, \hat{T})$,

$$|V_\alpha(w,\pi^*_\alpha) - V_\alpha(w,\hat{\pi})| \leq \frac{(1 - e^{-\lambda|T_\alpha - \hat{T}|})}{1 - \alpha} B \quad (3.221)$$

Now substituting (3.203) and (3.221) into (3.197), the bound

$$|\Phi(w,\hat{\pi}) - (1 - \alpha)\, V_\alpha(w,\pi^*_\alpha)| \leq (1 - \alpha^{N'})\,(\Phi(w,\hat{\pi}) + B_{N'}) + \alpha^{N'}\varepsilon$$
$$+ B(1 - e^{-\lambda|\hat{T} - T_\alpha|}) \quad . \quad (3.222)$$

Replacing $\alpha$ with $\alpha_n$ in (3.222), it is clear that for any $\varepsilon > 0$ there exists a N such that for all $n > N$

$$|\Phi(w,\hat{\pi}) - (1 - \alpha_n)\, V_{\alpha_n}(w,\pi^*_{\alpha_n})| \leq \varepsilon \quad (3.223)$$

Hence, from (3.223), the results of the lemma follow.

Q.E.D

Finally, using Theorem 3.2 the average optimal policy is established.

**Theorem 3.3**

For the SCADCP outlined in Section 3.2, which uses two data compressors with associated distortion levels $D_1 < D_2$ where $D_2 > r^{-1}(\frac{C}{\lambda} - p_2)$, the connected policy $\pi^*$ with threshold $\hat{T}$ given by (3.194) is the average optimal policy. The policy $\pi^*$ selects data compressor 1 when the waiting time, $w$, of a message satisfies $w \leq \hat{T}$, and selects data compressor 2 otherwise.

Proof

By Lemmas 3.3, 3.9, and 3.11, Assumptions 3.1, 3.2 and 3.3 are shown to be valid. Therefore, the results stated follow by Theorem 3.2.

Q.E.D

Theorem 3.3 provides that a stationary deterministic policy satisfies (3.31). This implies that a policy exists which satisfies (3.19). Thus, the representation for $\gamma_2^1(D)$ given in equation (3.16) is valid. Using the knowledge that the policy which satisfies (3.19) is a connected policy, $\gamma_2^1(D)$ is evaluated in Chapter IV.

## 3.5 Conclusions

In this chapter, the delay distortion relationship for communication networks was introduced and an adaptive decompression scheme was presented. The delay distortion relationship represents the minimum average message delay as a function of the fidelity of the reproduction of the sources. For the adaptive data compression scheme considered, a functional form of this relationship was presented. In the process of deriving the functional form, Markov decision processes were introduced and results pertaining to these processes were discussed. The results presented apply specifically to the case of Markov decision processes with nondenumerable state spaces, and unbounded cost functions. Using these results, for a single channel network the optimal structure of the adaptive data compression scheme was found for a class of adaptive compression schemes which employ two data compressors.

# CHAPTER IV

# DELAY DISTORTION RELATIONSHIP FOR A SINGLE CHANNEL

In this chapter the delay distortion relationship for a communication system employing a single channel is considered. The communication system utilizes an adaptive data compression scheme as described in Section 3.2. In Section 4.1, the results of Chapter III are incorporated into the evaluation of a delay distortion relationship. This relationship is specified by optimizing an appropriate functional. Some of the properties of this relationship are discussed in Section 4.2. In Section 4.3, the functional to be optimized is described in terms of an integral equation. This integral equation is solved for the case of a data compression scheme which uses two data compressors. For the case of a data compression scheme that utilizes more than two data compressors, an expression for the functional is derived in Section 4.4 for a dense set of values in the domain space of the functional. Using these results, numerical results for the delay distortion relationship for various schemes using multiple data compressors are presented in Section 4.5.

## 4.1 A Delay Distortion Relationship for a Single Channel

For the case of a single communication channel and a data compression scheme utilizing two data compressors, as described in Section 3.2, Theorem 3.3 proves that an optimal policy exists which

minimizes $L_\pi^2(\{D_k\})$, where $L_\pi^2(\{D_k\})$ is given by (3.15). In this section, using knowledge of the structure of the optimal policy, the delay distortion relationship is expressed as an optimization of an appropriate functional over a vector space.

The structure of the optimal policy is a connected policy $\pi^*$ with associated threshold $T^*$ applied to the observed waiting times. Thus, if $\Pi_c$ is the set of all stationary deterministic connected policies utilizing a threshold with respect to the observed waiting times, then $\pi^* \in \Pi_c$. So, in the expression for $L^2(\{D_k\})$, given by (3.18), the minimum can be taken over the set $\Pi_c$, giving

$$L^2(\{D_k\}) = \min_{\pi \in \Pi_c} L_\pi^2(\{D_k\}) \quad . \tag{4.1}$$

Since every $\pi \in \Pi_c$ is characterized by a threshold T, then clearly for all $\pi \in \Pi_c$, $L_\pi^2(\{D_k\})$ can be denoted $L_T^2(\{D_k\})$ where T is the threshold associated with $\pi$. So, using this result $L^2(\{D_k\})$ is given by

$$L^2(\{D_k\}) = \min_{T \geq -1} L_T^2(\{D_k\}) \tag{4.2}$$

where policies associated with negative values of T are the policies which never use the first data compressor whose distortion level is $D_1$. Now assuming the minimum over all $\{D_k \geq 0\}$ of $L^2(\{D_k\})$ exists, the delay distortion relationship $\gamma_2^1(D)$ given by (3.16) reduces to

$$\gamma_2^1(D^*) = \min_{\substack{\{D_k \geq 0, k=1,2\} \\ \{T \geq -1\}}} \left\{ L_T^2(\{D_k\}) \right\} - \mu D^* \tag{4.3}$$

where D* is the average distortion associated with $\{D_k^*\}$ and with the connected policy with threshold T* and $\{D_k^*\}$ and T* minimize $L_T^2(\{D_k\})$. Thus, the delay distortion relationship reduces to the minimization of a functional over a vector space.

When more than two data compressors are used in the compression scheme, $K > 2$, the structure of the optimal policy has not been determined. From the results in Chapter 3, it seems natural that the optimal policy should be a connected policy utilizing thresholds $\{T_k, k = 1, 2, \ldots, k-1\}$ with respect to the observed waiting times. This policy selects data compressor k for the $n^{th}$ message whenever the waiting time $w_n$ of the message satisfies $T_{k-1} < w_n \leq T_K$ where

$$T_0 = 0^- \tag{4.4}$$

and

$$T_K = \infty \quad . \tag{4.5}$$

So, restricting the class of data compression schemes to those employing stationary connected policies, a delay distortion relationship, to be denoted as $\tilde{\gamma}_K^{-1}(D)$, results. In order to determine $\tilde{\gamma}_K^{-1}(D)$, let $L_{\{T_k\}}^K(\{D_k\})$ be defined in a similar manner as $L_T^2(\{D_k\})$. So, if $\pi$ is a connected policy with thresholds $\{T_k\}$, then

$$L_{\{T_k\}}^K(\{D_k\}) = L_\pi^K(\{D_k\}) \quad . \tag{4.6}$$

Thus, the delay distortion relationship for connected policies is given by

$$\tilde{\gamma}_K^1(D^*) = \min_{\substack{\{D_k \geq 0, k=1,2,..,K\} \\ \{T_k \geq -1, k=1,2,...,K-1\}}} \left\{ L^K_{\{T_k\}}(\{D_k\}) \right\} - \mu D^* \qquad (4.7)$$

where $D^*$ is the average distortion associated with $\{D_k^*\}$ and with the connected policy with thresholds $\{T_k^*\}$ and $\{D_k^*\}$ and $\{T_k^*\}$ minimize $L^K_{\{T_k\}}(\{D_k\})$. Hence, the relationship $\tilde{\gamma}_K^1(D)$ reduces to the minimization of a functional over a vector space. Clearly, $\gamma_2^1(D)$ given in (4.3) is equal to $\tilde{\gamma}_2^1(D)$ and so, the remaining sections of this chapter are restricted to the discussion of $\tilde{\gamma}_K^1(D)$.

In order to evaluate $\tilde{\gamma}_K^1(D)$, expressions for $L^K_{\{T_k\}}(\{D_k\})$ must be derived. In Section 4.3 and 4.4, the expression $L^K_{\{T_k\}}(\{D_k\})$ is derived. Once these expressions are derived, optimization techniques are employed to determine $\tilde{\gamma}_K^1(D)$ in Section 4.5.

## 4.2 Properties of a Delay Distortion Relationship for a Single Channel

Some of the properties of the delay distortion relationship $\tilde{\gamma}_K^1(D)$ defined in Section 4.1 are outlined in this section. The first property is that $\tilde{\gamma}_K^1(D)$ is a nonincreasing function of the average distortion level D. This is clearly seen by considering (3.13), which gives

$$\tilde{\gamma}_K^1(D) = \inf_{\substack{\pi \epsilon \Pi(D;\{D_k\}) \\ \{D_k \geq 0, k=1,2,..,K\}}} T^K_\pi(\{D_k\}) \qquad (4.8)$$

where

$$\Pi(D;\ \{D_k\}) = \{\pi\ \epsilon\ \Pi_c:\ D^K_\pi(\{D_k\}) \leq D\} \qquad (4.9)$$

and $\Pi_c$ is the set of all stationary deterministic connected policies which utilize the observed waiting times as a statistic to base the decisions upon. Thus, for $\{D_k\}$ fixed, sets $\Pi(D; \{D_k\})$ over which the infimum is taken and are decreasing telescoping sets as the distortion level D increases (e.g., if $\tilde{D}_1 < \tilde{D}_2$ then $\Pi(\tilde{D}_1; \{D_k\}) \subseteq \Pi(\tilde{D}_2; \{D_k\}))$. Hence, the infimum in (4.8) is a non-increasing function of D.

Another property of $\tilde{\gamma}_K^1(D)$ which is of interest is the minimum value of distortion, $D_{min}^K$, such that $\tilde{\gamma}_K^1(D)$ is finite. So, $D_{min}^K$ is given by

$$D_{min}^K = \inf_{D \epsilon \mathcal{D}_K} D \tag{4.10}$$

where

$$\mathcal{D}_K = \{D: \tilde{\gamma}_K^1(D) \text{ is finite}\} . \tag{4.11}$$

For the case of K = 1, $D_{min}^1$ is specified in (3.8) to be $\hat{r}^{-1}(C/\lambda)$. For the case K > 1, the following proposition establishes the corresponding values of $D_{min}^K$.

__Proposition 4.1__

Let the rate distortion relationship, $\hat{r}(D)$, that the set of data compressors satisfy, be a decreasing convex function of D, and let $p_K$ be the protocol information required to specify to the decoder and the user which compressor is used to compress a message. Then $D_{min}^K$, defined in (4.10), is given by

$$D_{min}^K = \hat{r}^{-1}(C/\lambda - p_K) \quad . \tag{4.12}$$

Proof

Let $P_k^n$ be the average probability of using compressor k during the compression of n messages, where $P_k^n$ is given by

$$P_k^n = n^{-1} \sum_{i=0}^{n-1} \Pr(i^{th} \text{ message compressed by } k^{th} \text{ DC}) \quad . \tag{4.13}$$

Let

$$t_k = (\hat{r}(D_k) + p_K)C^{-1} \tag{4.14}$$

denote the transmission time of a message and the associated protocol information which use the $k^{th}$ data compressor. Clearly, the buffer in the communication system behaves as a queue in a queueing system with mean message interarrival time $\lambda^{-1}$ and average service time over n messages given by

$$\bar{\sigma}_n = \sum_{k=1}^{K} t_k P_k^n \tag{4.15}$$

Thus, the average traffic intensity, being the ratio of the average service time to mean interarrival time, is given by

$$\rho_n = \lambda \sum_{k=1}^{K} t_k P_k^n \quad . \tag{4.16}$$

Correspondingly, the average distortion is given by

$$D^n = \sum_{k=1}^{K} P_k^n D_k \quad . \tag{4.17}$$

Now assuming that $\{D_k\}$ and $P_k^n$ can be selected independently,

consider the problem of determining a set $\{D_k\}$ that minimizes $D^n$ for a fixed value of n subject to the constraint that the average traffic intensity is less than or equal to one, $\rho_n \leq 1$, for a given set $P_k^n$, $k = 1, 2, \ldots, K$. Using Jensen's inequality, it is clear that for every set $P_k^n$, $k = 1, \ldots, K$, the set $\{D_k\}$ that minimizes $D^n$ subject to the constraint is given for all k and n by

$$D_k = \hat{r}^{-1}(C/\lambda - p_K)$$

$$= D_K^* \tag{4.18}$$

To show that $D_{min}^K = D_K^*$, it is required to find schemes whose distortion levels are greater than $D_K^*$ which have finite limiting average delay and show that no scheme exists with finite limiting average delay with distortion less than or equal to $D_K^*$. First, let $D > D_K^*$ and $D_k = D$ for $k = 1, 2, \ldots, K$. Then, $D^n = D$ and $\rho_n < 1$ for all n, and the limiting average waiting time $\bar{W}$ for this M/D/1 queueing system is given by

$$\bar{W} = \frac{(\hat{r}(D) + p_K)C^{-1}}{2} \quad \frac{\lambda(\hat{r}(D) + p_K)C^{-1}}{1 - \lambda(\hat{r}(D) + p_K)C^{-1}} \tag{4.19}$$

Hence, from (4.18), the limiting average waiting time is finite, and so, the limiting average delay $\tilde{\gamma}_K^1(D)$ is finite for $D > D_K^*$.

For the case $D \leq D_K^*$, the waiting time of the $(n+1)^{st}$ message is given by (3.21) as

$$w_{n+1} = \max\{w_n + t_{a_i} - \tau_{n+1}, 0\}$$

$$= \max\left\{0,\ w_0 + \sum_{i=0}^{n} (t_{a_i} - \tau_{i+1}),\ \sum_{i=j}^{n} (t_{a_i} - \tau_{i+1}),\ j=1,2,..n\right\}. \tag{4.20}$$

So, $w_{n+1}$ is lower bounded by

$$w_{n+1} \geq \sum_{i=0}^{n}\left(t_{a_i} - \tau_{i+1}\right) \tag{4.21}$$

and using (4.16),

$$E(w_{n+1}) \geq n\,\lambda^{-1}(\rho_n - 1) \quad . \tag{4.22}$$

From Jensen's inequality if $D = \liminf_{n\to\infty} D^n \leq D_K^*$, then $\liminf_{n\to\infty} \rho_n \geq 1$. Thus, if $\liminf_{n\to\infty} \rho_n > 1$, from (4.22) $\liminf_{n\to\infty} E\{w_n\}$ is unbounded and the limiting average message delay is unbounded. For the case $\liminf_{n\to\infty} \rho_n = 1$ and $D \leq D_K^*$, D equals $D_K^*$ by (4.18) and the data compressors have distortion levels of $D_K^*$. Thus, from Lindley's theorem, the limiting average delay is unbounded. Hence, (4.12) is valid.

<u>Q.E.D</u>

## 4.3 Evaluation of $L^K_{\{T_k\}}(\{D_k\})$

In order to compute $L^K_{\{T_k\}}(\{D_k\})$ given by (4.6), an expression for the waiting time probability distribution F(W) must be obtained. In this section an integral equation is determined whose solution is F(W) and this integral equation is solved for the case of two data

compressors, $K = 2$.

Generalizing Lemma 3.10 to K data compressors, it is easily shown that if the data compressors are numbered in order of increasing distortion levels (e.g., $D_k \leq D_{k+1}$), with $t_K \leq \lambda^{-1}$ and $T_{K-1} < \infty$, then $F(W)$ is a valid distribution function. Since $F(W)$ exists, $L^K_{\{T_k\}}(\{D_k\})$ is given in terms of $F(W)$ by

$$L^K_{\{T_k\}}(\{D_k\}) = \lim_{n\to\infty} E_{\{T_k\}}(w_n + t_{a_n} + \mu D_{a_n})$$

$$= \int_0^\infty WF(dW) + \sum_{k=1}^{K} (t_k + \mu D_k)\,[F(T_k) - F(T_{k-1})] \qquad (4.23)$$

where $F(T_0) = 0$ and $F(T_K) = 1$. The following theorem specifies an integral equation whose solution is $F(W)$.

<u>Theorem 4.1</u>

The limit, $F(W)$, of the probability distribution of the waiting time of a message in a data compression scheme using K data compressors with distortion levels $\{D_k\}$, and using a connected stationary deterministic policy with thresholds $\{T_k\}$ satisfies, for $W \geq 0$,

$$F(W) = \sum_{k=1}^{K} \int_{\max(T_{k-1},\,W-t_k)}^{\infty} [F(\min(\tilde{W},T_k)) - F(T_{k-1})]\lambda\, e^{-\lambda(\tilde{W}+t_k-W)}\, d\tilde{W} \quad (4.24)$$

where $F(T_0) = 0$, $F(T_K) = 1$, $\min(w, T_K) = w$, and $t_k$ is given by (4.14). Furthermore, $F(W)$ is the unique solution to (4.24) subject to the constraint that $F(W)$ is a continuous and piecewise differentiable function and the

$$\lim_{W\to\infty} F(W) = 1 \quad . \qquad (4.25)$$

## Proof

Using the connected policy with thresholds $\{T_k\}$, it is clear that the waiting time of the $n^{th}$ message is given by

$$w_n = \max(0, w_{n-1} + t_k - \tau_n) \tag{4.26}$$

where k is such that $T_{k-1} < w_{n-1} \leq T_k$ and $\{\tau_n\}$ are independent identically distributed exponential random variables with mean $\lambda^{-1}$. Thus, $\Pr(w_n \leq W, T_{k-1} < w_{n-1} \leq T_k | \tau_n = \tau)$ is given by

$$\Pr(w_n \leq W, T_{k-1} < w_{n-1} \leq T_k | \tau_n = \tau) = \begin{cases} \Pr(T_{k-1} < w_{n-1} \leq T_k) & , \text{ if } T_k + t_k - \tau \leq W \\ \Pr(T_{k-1} < w_{n-1} \leq W - t_k + \tau), & \\ & \text{if } T_{k-1} + t_k - \tau \leq W < T_k + t_k - \tau \\ 0 & , \text{ if } T_{k-1} + t_k - \tau > W. \end{cases} \tag{4.27}$$

Averaging over $\tau$ in (4.27),

$$\begin{aligned} &\Pr(w_n \leq W, T_{k-1} < w_{n-1} \leq T_k) \\ &= \int_{\max(T_k + t_k - W, 0)}^{\infty} \Pr(T_{k-1} < w_{n-1} \leq T_k) \lambda e^{-\lambda\tau} d\tau \\ &\quad + \int_{\max(T_{k-1} + t_k - W, 0)} \Pr(T_{k-1} < w_{n-1} \leq W - t_k + \tau) \lambda e^{-\lambda\tau} d\tau \\ &= \int_{\max(T_{k-1} + t_k - W, 0)}^{\infty} \Pr(T_{k-1} < w_{n-1} \leq \min(T_k, W - t_k + \tau)) \lambda e^{-\lambda\tau} d\tau. \end{aligned} \tag{4.28}$$

Letting $\tilde{w} = \tau - t_k + W$, (4.28) becomes

$$\Pr(W_n \leq W,\ T_{k-1} < w_{n-1} \leq T_k)$$

$$= \int_{\max(T_{k-1},W-t_k)}^{\infty} \Pr(T_{k-1} < w_{n-1} \leq \min(T_k,\tilde{w}))\lambda\, e^{-\lambda(\tilde{w}+t_k-W)}\, d\tilde{w}\ . \quad (4.29)$$

Summing (4.29) over k, $\Pr(w_n \leq w)$ is given by

$$\Pr(w_n \leq W)$$

$$= \sum_{k=1}^{K} \int_{\max(T_{k-1},W-t_k)}^{\infty} \Pr(T_{k-1} < w_{n-1} \leq \min(T_k,\tilde{w}))\lambda\, e^{-\lambda(\tilde{w}+t_k-W)}\, d\tilde{w}\ . \quad (4.30)$$

Since the limiting probability distribution,

$$F(W) = \lim_{n\to\infty} \Pr(w_n \leq W)\ , \quad (4.31)$$

exists, the dominated convergence theorem gives (4.24).

Now suppose $F^*(W)$ is a continuous piecewise differentiable function which satisfies (4.24) and (4.25). The derivative of $F^*(W)$ is computed from (4.24) and is given by

$$\frac{dF^*(W)}{dW} = \lambda(F^*(W) - G^*(W)) \quad (4.32)$$

where

$$G^*(W) = \begin{cases} F^*(W - t_k)\ , & \text{if}\quad T_{k-1} < W - t_k \leq T_k,\ k = 1,2,\ldots,K \\ 0\ , & \text{otherwise}\ . \end{cases}$$

Since $F^*(W)$ is a continuous and piecewise differentiable function, $F^*(W)$ can be expressed as

$$F^*(W) = \int_0^W \left(\frac{dF(w)}{dw}\right) dw + F^*(0)\ . \quad (4.33)$$

Clearly from (4.32) and (4.33), if $F^*(0) \geq 0$, then $F^*(W)$ is a non-negative nondecreasing function; and if $F^*(0) < 0$, then $F^*(W)$ is a negative nonincreasing function. But $F^*(W)$ satisfies (4.25), and so $F^*(0) > 0$ and $F^*(W)$ is a valid distribution function. Now let $F^*(W)$ be the initial waiting time distribution of the queueing system and let the waiting time distribution of the $n^{th}$ message, $F^*_n(W)$, be given by

$$F_{n+1}(W) = \sum_{k=1}^{K} \int_{\max(T_{k-1},W-t_k)}^{\infty} [F^*_n(\min(\tilde{w},T_k)) - F^*_n(T_{k-1})]\lambda e^{-\lambda(\tilde{w}+t_k-W)} d\tilde{w}. \tag{4.34}$$

But from a generalization of Lemma 3.10

$$\lim_{n\to\infty} F^*_n(W) = F(W) . \tag{4.35}$$

Therefore, since $F^*(W)$ satisfies (4.24), it is clear that $F^*(W)$, $F(W)$ and $F_n(W)$ are all identical.

Q.E.D

For the case of more than two data compressors, $K > 2$, the solution to the integral equation (4.24) is quite complex. For this case, an alternative expression for $L^K_{\{T_k\}}(\{D_k\})$ is discussed in Section 4.4. The following theorem determines $F(W)$ for the case $K = 2$.

Theorem 4.2

Let $\mu_i = e^{-\lambda t_i}$, $i = 1, 2$. Then the solution of the integral equation (4.24) for $K = 2$ is given by

$$F(W) = \begin{cases} c\, e^{\lambda W}\, G^{0}_{t_1}(W) & , \text{ if } \quad 0 \le W \le T_1 \\ c\, e^{\lambda W}\Big\{G_T\, \mu_2\, \lambda^{-1}[G^{T_1+t_2}_{t_2}(W) - G^{T_1+t_1}_{t_2}(W)] + \mu_2\, G_2\, G^{T_1+t_1}_{t_2}(W) & \\ \quad - G_T[e^{-\lambda t_2}\, G_e(W - T_1 - t_2) - e^{-\lambda t_1}\, G_e(W - T_1 - t_1)] & \\ \quad + [G^{T_1}_{G}(W) - G^{T_1+t_1}_{G}(W)]\Big\} & , \text{ if } \quad T_1 < W \end{cases} \tag{4.36}$$

for the case $T_1 + t_2 \le t_1$ and is given by

$$F(W) = \begin{cases} c\, e^{\lambda W}\, G^{t_1}_{0}(W) & , \text{ if } 0 \le W \le T_1 \\ c\, e^{\lambda W}\Big\{[G^{T_1}_{t_2}(W) - G^{T_1}_{t_2}(W)] + G_T\, \mu_2\, \lambda^{-1}[G^{T_1+t_2}_{t_2}(W) - G^{T_1+t_1}_{t_2}(W)] & \\ \quad - G_{T_1}[e^{-\lambda t_2}\, G_e(W - T_1 - t_2) - e^{-\lambda t_1}\, G_e(W - T_1 - t_1)] & \\ \quad + [G^{t_1}_{G}(W) - G^{T_1+t_1}_{G}(W)] + \mu_2\, G_2\, G^{T_1+t_1}_{t_2}(W)\Big\} & , \text{ if } T_1 < W \end{cases} \tag{4.37}$$

for the case $t_1 < T_1 + t_2$. The constants $G_1$, $G_2$, $G_T$ and c in (4.36) and (4.37) are given by

$$G_1 = (1 - G^{0}_{t_0}(T_1 + t_1))\mu_1^{-1} \quad , \tag{4.38}$$

$$G_2 = \{G^{0}_{t_1}(T_1 + t_1) - G^{0}_{t_1}(T)\,(\mu_1 - \mu_2)\lambda^{-1}\}\mu_2^{-1} \quad , \tag{4.39}$$

$$G_T = G^0_{t_1}(T) \tag{4.40}$$

and

$$c^{-1} = \begin{cases} \dfrac{G_T\,\mu_2\,\lambda^{-1}}{1-\lambda t_2}\left\{e^{\lambda(T_1+t_2)} - e^{\lambda(T_1+t_1)}\right\} + \dfrac{\mu_2\,G_2}{(1-\lambda t_2)}\,e^{\lambda(T_1+t_1)} \\ \quad + \dfrac{G_T\,e^{\lambda T}(t_1-t_2)}{(1-\lambda t_2)} + \displaystyle\sum_{r=0}^{\left[\frac{T_1+t_1}{t_1}\right]} (L_r^{T_1} - L_r^{T+t_1}) & \text{, if } T_1 + t_2 \le t_1 \\ \dfrac{e^{\lambda T_1} - e^{\lambda t_1}}{(1-\lambda t_2)} + \dfrac{G_T\,\mu_2\,\lambda^{-1}}{(1-\lambda t_2)}\left\{e^{\lambda(T_1+t_2)} - e^{\lambda(T_1+t_1)}\right\} \\ \quad + \dfrac{\mu_2\,G_2}{(1-\lambda t_2)}\,e^{\lambda(T_1+t_1)} + \dfrac{G_T\,e^{\lambda T_1}(t_1-t_2)}{(1-\lambda t_2)} \\ \quad + \displaystyle\sum_{r=0}^{\left[\frac{T_1+t_1}{t_1}\right]} (L_r^{t_1} - L_r^{T_1+t_1}) & \text{, if } t_1 < T_1 + t_2 \end{cases} \tag{4.41}$$

The symbol [x] is the greatest integer less than or equal to x and $L_r^T$ is given by (A.53) in the Appendix. The functions $G^T_{t_1}(W)$, $G^T_G(W)$ and $G_e(W)$ are given in (A.16), (A.24), and (A.39), respectively, in the Appendix.

**Proof**

Let

$$G(W) = c^{-1}\,e^{-\lambda W}\,F(W) \tag{4.42}$$

where c is selected such that

$$G(0) = 1 \quad . \tag{4.43}$$

So, upon substituting of (4.42) into (4.24) the resulting integral equation for G(W) is given by

$$G(W) = \sum_{k=1}^{2} \mu_k \int_{\max(T_{k-1}, W-t_k)}^{\infty} \left[ G(\min(\tilde{w}, T_k)\, e^{\lambda(\min(0, T_k - \tilde{w})} - G(T_{k-1})\, e^{\lambda(T_{k-1} - \tilde{w})} \right] d\tilde{w} \quad . \tag{4.44}$$

So, simplifying (4.44) and letting $T = T_1$,

$$G(W) = \begin{cases} \mu_1 \int_0^T G(y)\, dy + \mu_2 \int_T^{\infty} G(y)\, dy + G(T)\, (\mu_1 - \mu_2)\lambda^{-1} \ , & \text{if} \quad W \le \min(t_1,\ T + t_2) \\ \mu_1 \int_{W-t_1}^{T} G(y)\, dy + \mu_2 \int_T^{\infty} G(y)\, dy + G(t)\, (\mu_1 - \mu_2)\lambda^{-1} \ , & \text{if} \quad t_1 < W \le T + t_2 \\ \mu_1 \int_0^T G(y)\, dy + \mu_2 \int_{W-t_2}^{\infty} G(y)\, dy + G(T)\, (\mu_1\, \lambda^{-1} - e^{-\lambda(W-T)}) \ , & \text{if} \quad T + t_2 < W \le t_1 \\ \mu_1 \int_{W-t_1}^{T} G(y)\, dy + \mu_2 \int_{W-t_2}^{\infty} G(y)\, dy + G(T)\, (\mu_1\, \lambda^{-1} - e^{-\lambda(W-T)}), & \text{if} \quad \max(t_1,\ T + t_2) < W \le T + t_1 \\ \mu_2 \int_{W-t_2}^{\infty} G(y)\, dy \ , & \text{if} \quad T + t_1 < W \quad . \end{cases} \tag{4.45}$$

Now let

$$G_1 = \int_0^T G(y)\, dy \tag{4.46}$$

$$G_2 = \int_T^\infty G(y)\, dy \tag{4.47}$$

$$G_T = G(T) \tag{4.48}$$

and

$$G_0 = \mu_1 G_1 + \mu_2 G_2 + G_T(\mu_1 - \mu_2)\lambda^{-1} \quad . \tag{4.49}$$

From (4.45)

$$G(0) = G_0 \quad , \tag{4.50}$$

and so from (4.43)

$$G_0 = 1 \quad . \tag{4.51}$$

Thus, upon substituting (4.46) - (4.51) into (4.45), the expression for G(W) reduces to

$$
G(W) = \begin{cases}
1 & , \quad \text{if} \quad W \le \min(t_1, T + t_2) \\
1 - \mu_1 \int_0^{W-t_1} G(y)\, dy & , \quad \text{if} \quad t_1 < W \le T + t_2 \\
1 + G_T(\mu_2\, \lambda^{-1} - e^{-\lambda(W-T)}) - \mu_2 \int_T^{W-t_1} G(y)\, dy \ , & \\
 & , \quad \text{if} \quad T + t_2 < W \le t_1 \\
1 + G_T(\mu_2\, \lambda^{-1} - e^{-\lambda(W-T)}) - \mu_1 \int_0^{W-t_1} G(y)\, dy & \\
\quad - \mu_2 \int_T^{W-t_2} G(y)\, dy, \text{ if} & \max(t_1, T + t_2) < W \le T + t_1 \\
\mu_2 G_2 - \mu_2 \int_T^{W-t_2} G(y)\, dy & , \quad \text{if} \quad T + t_1 < W \ .
\end{cases}
\tag{4.52}
$$

In solving (4.52) for G(W), two cases need to be considered. These are $t_1 \le T + t_2$ and $t_1 > T + t_2$. For the first case of $t_1 \le T + t_2$, the solution is found by dividing the domain of G(W) into regions and solving for G(W) in these regions. From (4.52), in the region $0 \le W \le t_2$

$$
G(W) = 1 \quad , \tag{4.53}
$$

and in the region $t_1 < W \le T + t_2$, G(W) *satisfies the* integral equation

$$
G(W) = 1 - \mu_1 \int_0^{W-t_1} G(y)\, dy \quad . \tag{4.54}
$$

Letting G(W) = 0 for W < 0, it is clear that (4.54) is in the form of (A.15) for $0 \le W \le T + t_2$. The solution to (A.15) is given in Lemma A.2 in the Appendix. Hence, G(W) is given by

$$G(W) = G^0_{t_1}(W) \tag{4.55}$$

for $0 \leq W \leq T + t_2$ where $G^0_{t_1}(W)$ is given by (A.16). From (4.52), in the region $T + t_2 < W \leq T + t_1$, $G(W)$ satisfies

$$G(W) = 1 + G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)}) - \mu_1 \int_0^{W-t_1} G(y)\, dy - \mu_2 \int_0^{W-t_2} G(y)\, dy\ . \tag{4.56}$$

Since $W - t_1 < T + t_2$, the expression for $G(W)$, (4.55), can be used to evaluate $\mu_1 \int_0^{W-t_1} G(y)\, dy$ to get

$$\mu_1 \int_0^{W-t_1} G(y)\, dy = \mu_1 \int_0^{W-t_1} G^0_{t_1}(y)\, dy \quad . \tag{4.57}$$

Using (A.15), (4.57) *becomes*

$$\mu_1 \int_0^{W-t_1} G(y)\, dy = - G^0_{t_1}(W) + 1 \tag{4.58}$$

and upon substitution in (4.56), $G(W)$ satisfies

$$G(W) = G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)}) + G^0_{t_1}(W) - \mu_2 \int_0^{W-t_2} G(y)\, dy\ . \tag{4.59}$$

Finally, from (4.52) in the region $T + t_1 < W$, $G(W)$ satisfies

$$G(W) = \mu_2\, G_2 - \mu_2 \int_0^{W-t_2} G(y)\, dy\ . \tag{4.60}$$

Combining (4.54), (4.59) and (4.60), for $W \geq 0$, $G(W)$ satisfies

$$G(W) = G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)})\, u(W - T - t_2) + G_{t_1}^0(W)\, u(W)$$
$$+ \{\mu_2 G_2 - G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)}) - G_{t_1}^0(W)\}\, u(W - T - t_1)$$
$$- [\mu_2 \int_0^{W-t_2} G(y)\, dy]\, u(W - T - t_2) \tag{4.61}$$

where

$$u(W) = \begin{cases} 1 & W \geq 0 \\ 0 & W < 0 \end{cases} \quad . \tag{4.62}$$

Clearly for $W \geq T$, $G(W)$ satisfies an integral equation of the form (A.1). Using the solution (A.5) derived in the Appendix, $G(W)$ is given for the case $t_1 \leq T + t_2$ by

$$G(W) = \begin{cases} G_{t_1}^0(W) \quad , & \text{if} \quad 0 \leq W \leq T \\ G_T\, \mu_2\lambda^{-1}\{G_{t_2}^{T+t_2}(W) - G_{t_2}^{T+t_1}(W)\} + \mu_2\, G_2\, G_{t_2}^{T+t_1}(W) & \\ \quad - G_T\{e^{-\lambda t_2} G_e(W - T - t_2) - e^{-\lambda t_1} G_e(W - T - t_1)\} & \\ \quad + \{G_G^T(W) - G_G^{T+t_1}(W)\} \quad , & \text{if} \quad T < W \end{cases} \tag{4.63}$$

where $G_{t_1}^T(W)$ satisfies (A.15) and is given by (A.16), $G_G^T(W)$ satisfies (A.23) and is given by (A.24), and $G_e(W)$ satisfies (A.38) and is given by (A.39).

For the second case of $t_1 > T + t_2$, again the solution is found

by dividing the domain of $G(W)$ into regions and solving for $G(W)$ in these regions. From (4.52), in the region $0 \le W \le T + t_2$

$$G(W) = 1 = G^0_{t_1}(W) \tag{4.64}$$

where $G^0_{t_1}(W)$ is given by (A.16). From (4.52), in the region $T + t_2 < W \le t_1$, $G(W)$ satisfies

$$G(W) = 1 + G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)}) - \mu_2 \int_0^{W-t_2} G(y)\, dy \tag{4.65}$$

and in the region $t_1 < W \le T + t_1$, $G(W)$ satisfies

$$G(W) = 1 + G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)}) - \mu_1 \int_0^{W-t_1} G(y)\, dy - \mu_2 \int_0^{W-t_2} G(y)\, dy\ . \tag{4.66}$$

Now for $W - t_1 \le T + t_2$, the expression for $G(W)$, (4.64), can be used to evaluate $\mu_1 \int_0^{W-t_1} G(y)\, dy$ to get

$$\mu_1 \int_0^{W-t_1} G(y)\, dy = \mu_1 \int_0^{W-t_1} G^0_{t_1}(y)\, dy\ . \tag{4.67}$$

Using (A.15) in (4.67), the expression in (4.67) becomes

$$\mu_1 \int_0^{W-t_1} G(y)\, dy = - G^0_{t_1}(W) + 1 \quad . \tag{4.68}$$

Upon substitution of (4.68) into (4.66), $G(W)$ in the region $t_1 < W \le T + t_1$ satisfies

$$G(W) = G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)}) + G^0_{t_1} - \mu_2 \int_0^{W-t_2} G(y)\, dy \quad . \tag{4.69}$$

Finally, from (4.52), in the region $T + t_1 < W$, $G(W)$ satisfies

$$G(W) = \mu_2 G_2 - \mu_2 \int_0^{W-t_2} G(y)\, dy \quad . \tag{4.70}$$

Combining (4.64), (4.65), (4.69) and (4.70), for $W \geq 0$, $G(W)$ satisfies

$$\begin{aligned} G(W) = u(W) &+ G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)})\, u(W - T - t_2) + (G^0_{t_1} - 1)\, u(W - t_1) \\ &+ \{\mu_2 G_2 - G_T(\mu_2\lambda^{-1} - e^{-\lambda(W-T)}) - G^0_{t_1}\}\, u(W - T - t_1) \\ &- [\mu_2 \int_0^{W-t_2} G(y)\, dy]\, u(W - T - t_2) \quad . \end{aligned} \tag{4.71}$$

As in the first case, for $T \leq W$, $G(W)$ clearly satisfies an integral equation of the form (A.1). Using the solution (A.5) derived in the Appendix, $G(W)$ is given for the case $t_1 > T + t_2$ by

$$G(W) = \begin{cases} G^0_{t_1}(W) & , \text{ if } \; 0 \leq W \leq T \\ \{G^T_{t_2}(W) - G^{t_1}_{t_2}(W)\} + G_T\, \mu_2\lambda^{-1}\{G^{T+t_2}_{t_2}(W) - G^{T+t_1}_{t_2}(W)\} & \\ \quad - G_T\{e^{-\lambda t_2} G_e(W - T - t_2) - e^{-\lambda t_1} G_e(W - T - t_1)\} & \\ \quad + \{G^{t_1}_G(W) - G^{T+t_1}_G(W)\} + \mu_2 G_2 G^{T+t_1}_{t_2}(W) & , \text{ if } \; T < W \quad . \end{cases} \tag{4.72}$$

In order to find $F(W)$, the constants $G_1$, $G_2$, and $G_T$ must be

evaluated. From (4.48), (4.63), and (4.72) $G_T$ is given by

$$G_T = G^0_{t_1}(T) \quad . \tag{4.73}$$

From (4.46), (4.63) and (4.72), $G_1$ is given by

$$G_1 = \int_0^T G(y)\,dy$$

$$= \int_0^T G^0_{t_1}(y)\,dy \quad . \tag{4.74}$$

Using (A.15) in (4.74), the expression for $G_1$ becomes

$$G_1 = (1 - G^0_{t_1}(T + t_1))\mu_1^{-1} \quad . \tag{4.75}$$

Substituting (4.73) and (4.75) into (4.47) and solving for $G_2$, $G_2$ is expressed by

$$G_2 = \{G^0_{t_1}(T + t_1) - G^0_{t_1}(T)\,(\mu_1 - \mu_2)\lambda^{-1}\}\mu_2^{-1} \quad . \tag{4.76}$$

Finally the constant c in equation (4.42) must be evaluated. It is evaluated using the constraint $\lim_{W\to\infty} F(W) = 1$ given in (4.25). From (4.42), this constraint implies that c satisfies

$$c^{-1} = \lim_{W\to\infty} (e^{\lambda W} G(W)) \quad . \tag{4.77}$$

Now using the expressions (4.63) and (4.72) for G(W) in (4.77) and then making use of the expressions (A.45), (A.52), and (A.62), the expression (4.41) for $c^{-1}$ is obtained.

Hence, solving (4.42) for F(W) and using (4.41), (4.63), and (4.72), the expressions (4.36) and (4.37) for F(W) are verified.

Q.E.D

Using the expressions (4.36) and (4.37) for F(W), the limiting average waiting time is evaluated in the following theorem.

Theorem 4.3

The mean,

$$E(W) = \int_0^\infty W d\,F(W) \quad , \tag{4.78}$$

associated with F(W) is given by

$$E(W) = c\Bigg\{-\frac{1}{\lambda}\sum_{r=0}^{[\frac{T_1}{t_1}]}\Big(\sum_{u=0}^{r} e^{\lambda(T_1-rt_1)}\Big(\frac{-\lambda(T_1-rt_1)^u}{u!}-1\Big)$$

$$+ G_T\,\mu_2\lambda^{-1}\Bigg[e^{\lambda(T_1+t_2)}\Big(\frac{T_1+t_2}{1-\lambda t_2}+\frac{\lambda t_2^{\,2}}{2(1-\lambda t_2)^2}\Big)$$

$$- e^{\lambda(T_1+t_1)}\Big(\frac{T_1+t_1}{1-\lambda t_2}+\frac{\lambda t_2^{\,2}}{2(1-\lambda t_1)^2}\Big)\Bigg]$$

$$+ \mu_2\,G_2\Big(\frac{T_1+t_1}{1-\lambda t_2}+\frac{\lambda t_2^{\,2}}{2(1-\lambda t_2)^2}\Big)$$

$$+ G_T\,e^{\lambda T_1}\Bigg(\frac{\lambda(T_1+t_1)^2-\lambda(T_1+t_2)^2}{2(1-\lambda t_2}+\frac{t_1-t_2}{1-\lambda t_2}+\frac{(\lambda t_2)^2(t_1-t_2)}{2(1-\lambda t_2}\Bigg)$$

$$+\frac{1}{1-\lambda t_2}\sum_{r=0}^{[\frac{T_1+t_1}{t_1}]}\big(h_r^{T_1}-h_r^{T_1+t_1}\big)\Bigg\} \tag{4.79}$$

for $T_1 + t_2 \leq t_1$ and is given by

$$E(W) = c\left\{ -\frac{1}{\lambda} \sum_{r=0}^{[\frac{T_1}{t_1}]} \left( \sum_{u=0}^{r} e^{\lambda(T_1 - rt_1)} \frac{(-\lambda(T_1 - rt_1))^u}{u!} - 1 \right) \right.$$

$$+ G_T \mu_2 \lambda^{-1} \left[ e^{\lambda T_1} \left( \frac{T_1}{1 - \lambda t_2} + \frac{\lambda t_2^{\,2}}{2(1 - \lambda t_2)^2} \right) \right.$$

$$\left. - e^{\lambda t_1} \left( \frac{t_1}{1 - \lambda t_2} + \frac{\lambda t_2^{\,2}}{2(1 - \lambda t_2)^2} \right) \right]$$

$$+ G_T \mu_2 \left[ e^{\lambda(T_1 + t_2)} \left( \frac{T_1 + t_2}{1 - \lambda t_2} + \frac{\lambda t_2^{\,2}}{2(1 - \lambda t_2)^2} \right) \right.$$

$$\left. - e^{\lambda(T_1 + t_1)} \left( \frac{T_1 + t_1}{1 - \lambda t_2} + \frac{\lambda t_2^{\,2}}{2(1 - \lambda t_2)^2} \right) \right]$$

$$+ \mu_2 G_2 \left( \frac{T_1 + t_1}{1 - \lambda t_2} + \frac{\lambda t_2^{\,2}}{2(1 - \lambda t_2)^2} \right)$$

$$+ G_T e^{\lambda T_1} \left( \frac{\lambda(T_1 + t_1)^2 - \lambda(T_1 + t_2)^2}{2(1 - \lambda t_2)} + \frac{t_1 - t_2}{1 - \lambda t_2} + \frac{(\lambda t_2)^2 (t_1 - t_2)}{2(1 - \lambda t_2)} \right)$$

$$\left. + \frac{1}{1 - \lambda t_2} \sum_{r=0}^{[\frac{T_1 + t_1}{t_1}]} \left( h_r^{t_1} - h_r^{T_1 + t_1} \right) \right\} \qquad (4.80)$$

for $t_1 < T_1 + t_2$ where $G_1$, $G_2$, $G_T$ and $c$ are given by (4.38), (4.39), (4.40), and (4.41), respectively and $h_r^{T_1}$ is given by (A.72) in the Appendix.

**Proof**

The mean associated with $F(W)$ is given by

$$E(W) = \lim_{W\to\infty} \int_0^W y \, dF(y) \quad . \tag{4.81}$$

Let $u(x)$ be the unit step function defined by

$$u(W) = \begin{cases} 1 & , \text{ if } W \geq 0 \\ 0 & , \text{ if } W < 0 \end{cases} \quad . \tag{4.82}$$

Then, from (4.81) $E(W)$ can be expressed by

$$E(W) = \int_0^{T_1} y \, dF(y) - T_1 \, F(T_1) + \lim_{W\to\infty} \int_0^W y \, d[u(y - T_1) \, F(y)] \ . \tag{4.83}$$

Integrating by parts, $E(W)$ is expressed by

$$E(W) = - \int_0^{T_1} F(y) \, dy + \lim_{W\to\infty} \int_0^W y \, d(u(y - T_1) \, F(y)) \quad . \tag{4.84}$$

From (4.36) and (4.37), for $0 \leq W \leq T$, $F(W)$ is given by

$$F(W) = c \, G^0_{t_1}(W) \, e^{\lambda W} \tag{4.85}$$

So, using the expression for $G^0_{t_1}(W)$, (A.16), the first term of (4.84) is given by

$$\int_0^{T_1} F(y)\,dy = c\int_0^{T_1} \sum_{r=0}^{[y/t_1]} \frac{(-\lambda(y - rt_1))^r}{r!} e^{\lambda(y-rt_1)}\,dy$$

$$= \frac{c}{\lambda} \sum_{r=0}^{[T_1/t_1]} \left(\sum_{u=0}^{r} e^{\lambda(T_1-rt_1)} \frac{(-\lambda(T_1 - rt_1))^u}{u!} - 1\right) . \tag{4.86}$$

Now, define the functions $H_{t_2}^{T}(x)$, $H_G^{T_1T_2}(x)$ and $H_e^{T_1T_2}(x)$ by

$$H_{t_2}^{T}(x) = \int_0^x y\, d(e^{\lambda y} G_{t_2}^{T}(y)) \tag{4.87}$$

$$H_G^{T_1T_2}(x) = \int_0^x y\, d(e^{\lambda y}\{G_G^{T_1}(y) - G_G^{T_2}(y)\}) \tag{4.88}$$

and

$$H_e^{T_1T_2}(x) = \int_0^x y\, d(e^{\lambda x}\{G_e(x - T_1) - e^{-\lambda(T_2-T_1)} G_e(x - T_2)\}) . \tag{4.89}$$

So, substituting the expression for $F(W)$ from (4.36) and (4.37) into (4.84) and using (4.87), (4.88) and (4.89), the expression for $E(W)$ becomes

$$E(W) = c\Bigg\{-\frac{1}{\lambda} \sum_{r=0}^{[T_1/t_1]} \left(\sum_{u=0}^{r} e^{\lambda(T_1-rt_1)} \frac{(-\lambda(T_1 - rt_1))^u}{u!} - 1\right)$$

$$+ G_T\, \mu_2 \lambda^{-1}[\lim_{x\to\infty} H_{t_2}^{T_1+t_2}(x) - \lim_{x\to\infty} H_{t_2}^{T_1+t_1}(x)] + \mu_2 G_2 \lim_{x\to\infty} H_{t_2}^{T_1+t_1}(x)$$

$$- G_T\, e^{\lambda T_1} \lim_{x\to\infty} H_e^{T_1+t_2,T_1+t_1} + \lim_{x\to\infty} H_G^{T_1,T_1+t_1}(x)\Bigg\} \tag{4.90}$$

for $T_1 + t_2 \leq t_1$ and becomes

$$
\begin{aligned}
E(\hat{W}) = c\Bigg\{ & -\frac{1}{\lambda} \sum_{r=0}^{[T_1/t_1]} \Big( \sum_{u=0}^{r} e^{\lambda(T_1 - rt_1)} \frac{(-\lambda(T_1 - rt_1))^u}{u!} - 1\Big) \\
& + \lim_{x\to\infty} H_{t_2}^{T_1}(x) - \lim_{x\to\infty} H_{t_2}^{t_1}(x) + G_T\, \mu_2 \lambda^{-1}\Big[\lim_{x\to\infty} H_{t_2}^{T_1+t_2}(x) \\
& - \lim_{x\to\infty} H_{t_2}^{T_1+t_1}(x)\Big] - G_T\, e^{\lambda T_1} \lim_{x\to\infty} H_e^{T_1+t_2, T_1+t_1}(x) \\
& + \lim_{x\to\infty} H_G^{t_1, T_1+t_1}(x) + \mu_2\, G_2 \lim_{x\to\infty} H_{t_2}^{T_1+t_1}(x)\Bigg\}
\end{aligned}
\tag{4.91}
$$

for $t_1 < T_1 + t_2$. The limit of the functions $H_{t_2}^{T}(x)$, $H_G^{T_1,T_2}(x)$ and $H_e^{T_1 T_2}(x)$ are evaluated in the Appendix in (A.67), (A.71), and (A.76). Hence, upon substituting these limiting values into (4.90) and (4.91), the results of the theorem are obtained.

<u>Q.E.D</u>

So, using the expressions derived for $F(W)$ and $\int_0^\infty F(W)\,dW$ in (4.23), $L^2_{\{T_k\}}(\{D_k\})$ is evaluated. Figure 4.1 summarizes this computation. In Section 4.5, these techniques for the evaluation of $L^2_{\{T_k\}}(\{D_k\})$ are used to obtain numerical results for $\gamma_2^1(D)$.

## 4.4 Approximation of the Delay Distortion Relationship

In the previous section, the integral equation (4.24) for the limiting probability distribution, $F(W)$, of the waiting time of a message in a data compression scheme using two data compressors ($K=2$) is solved explicitly for $F(W)$. This allows $L^2_{\{T_k\}}(\{D_k\})$ to be evaluated.

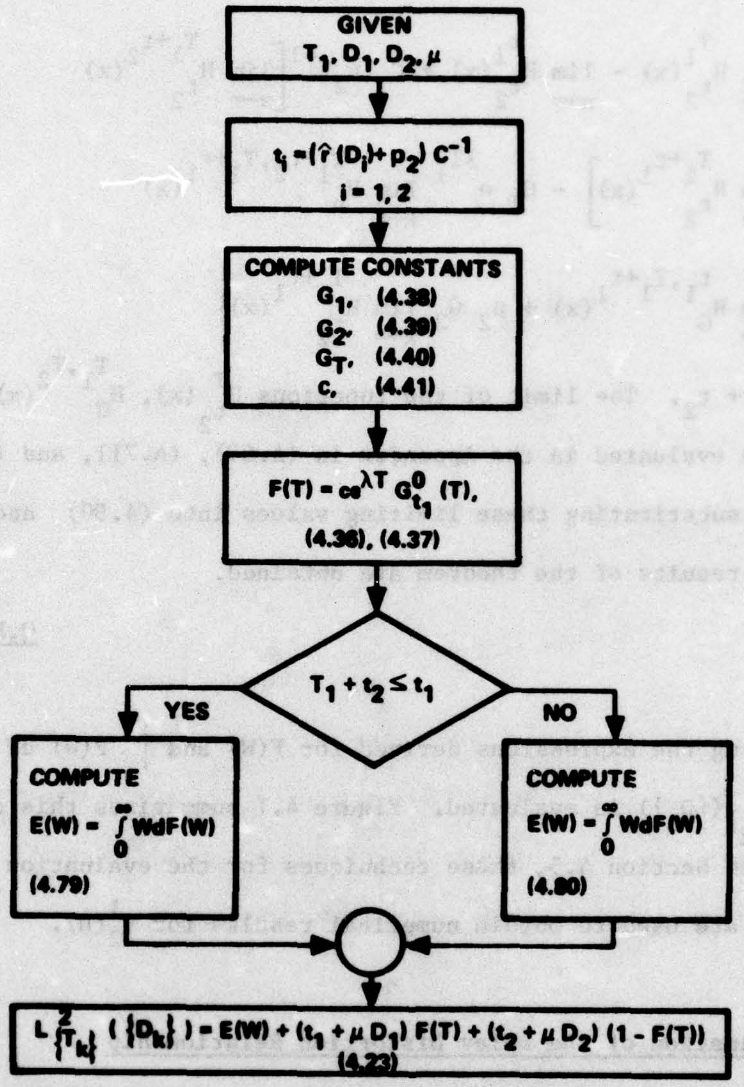



Figure 4.1. Evaluation of $L^2_{\{T_k\}}(\{D_k\})$

For K greater than two the corresponding integral equation (4.24) can be solved by the usual numerical techniques for solving integral equations. Then $L^{K}_{\{T_k\}}(\{D_k\})$ can be evaluated and $\tilde{\gamma}^{1}_{K}(D)$ can be computed using (4.7).

In this section, an alternative technique for evaluating $L^{K}_{\{T_k\}}(\{D_k\})$ is presented which does not rely on numerical techniques and their inherent convergence difficulties to solve for F(W). The technique evaluates $L^{K}_{\{T_k\}}(\{D_k\})$ by considering an appropriate imbedded Markov chains when the squences $\{T_k,\ k = 1, \ldots, K-1\}$ and $\{t_k,\ k = 1, 2, \ldots, K\}$ have a common divisor denoted by $\Delta$ (i.e., dividing each term of the sequences by $\Delta$ yields an integer). Then $\tilde{\gamma}^{1}_{K}(D)$ is evaluated using (4.7) where the minimum operation is taken over a dense set of values of $\{D_k,\ k = 1, 2, \ldots, K\}$ and a dense set of values of $\{T_k,\ k = 1, 2, \ldots, K-1\}$ (i.e., the corresponding set of values $\{t_k,\ k = 1, \ldots, K\}$ and the set of values $\{T_k,\ k = 1, \ldots, K-1\}$ have a common divisor).

In order to define the Markov chains used in the evaluation of $L^{K}_{\{T_k\}}(\{D_k\})$, define the integers $\{\ell_k,\ k = 1, 2, \ldots, K\}$ and $\{m_k,\ m = 1, 2, \ldots, K-1\}$ by

$$\ell_k = \frac{t_k}{\Delta} \qquad k = 1, 2, \ldots, K \tag{4.92}$$

and

$$m_k = \frac{T_k}{\Delta} \qquad k = 1, 2, \ldots, K-1\ . \tag{4.93}$$

Then for the queueing system with thresholds $\{T_k\}$ and service time $\{t_k\}$, the virtual waiting time at an instant $n\Delta$ units of time following

the initiation of a busy period, is an integer multiple of $\Delta$. So, define the sequence of imbedded times $\{\hat{t}_n\}$, $n \geq 1$, by

$$\hat{t}_{n+1} = \begin{cases} \hat{t}_n + \Delta & \text{, if } \hat{t}_n \text{ is not the end of a busy period} \\ \max\{t_{b_j} : t_{b_j} > \hat{t}_n, j \geq 0\}, & \text{otherwise} \end{cases} \tag{4.94}$$

where $\hat{t}_1 = t_{b_1}$, and $t_{b_j}$ is the time at which the $j^{th}$ busy period begins. Thus, if $V(t)$ is the virtual waiting time of the queueing system with

$$V(0) = 0 \quad , \tag{4.95}$$

then the sequence $\{x_n\}$ defined by

$$x_n = V(\hat{t}_n)/\Delta \qquad n = 1, 2, \ldots \tag{4.96}$$

is an imbedded Markov chain.

Clearly, $x_n$ can be written as a function of $x_{n-1}$ and the number of messages $N_n$ arriving in the interval $(\hat{t}_{n-1}, \hat{t}_n]$. Hence, if $x_{n-1} = 0$, then $\hat{t}_n$ is the beginning of a busy period and so, $N_n = 1$ and $x_n = \ell_1$. When $x_{n-1} > 0$, the virtual waiting $V(\hat{t}_n)$ is given by

$$V(\hat{t}_n) = V(\hat{t}_{n-1}) - \Delta + S_n \tag{4.97}$$

where $S_n$ is the sum of the transmission times of messages arriving in the interval $(\hat{t}_{n-1}, \hat{t}_n]$. To compute $V(\hat{t}_n)$ or equivalently $x_n$, let $Z(N,i)$ be the value of $x_n$ when $x_{n-1} = i > 0$, and $N$ messages arrived in the interval $(\hat{t}_{n-1}, \hat{t}_n]$. So, using the normalized thresholds $\{m_k\}$,

$Z(N,i)$ can be computed recursively by

$$Z(N,i) = Z(N-1,i) + \ell_k \quad , \text{ if } \; m_{k-1} < Z(N-1,i) - 1 \leq m_k, \; k = 1, \ldots, K \tag{4.98}$$

where $Z(0,i) = i-1$. Therefore for $x_{n-1} > 0$,

$$x_n = Z(N_n, x_{n-1}) \tag{4.99}$$

which is computed using (4.98).

To compute the transition probability

$$P_{j|i} = Pr(x_n = j | x_{n-1} = i) \quad , \tag{4.100}$$

first note that $Pr(N_n = N | x_{n-1} > 0)$, denoted by $P(N)$, is given by a Poisson distribution, and so

$$Pr(N_n = N | x_{n-1} > 0) = \frac{(\lambda\Delta)^N}{N!} e^{-\lambda\Delta}. \tag{4.101}$$

Hence,

$$P_{j|i} = \begin{cases} 1 & , \text{ if } \; j = \ell_1, \; i = 0 \\ P(N) & , \text{ if } \; j = Z(N,i), \; i > 0, \; N = 0, 1, 2, 3, \ldots \\ 0 & , \text{ otherwise} \end{cases} \quad . \tag{4.102}$$

Now if $T_{K-1} < \infty$ and $\lambda t_k < 1$, then it can be shown that the imbedded Markov chain $\{x_n\}$ is aperiodic and positive recurrent. Thus, $\lim_{n\to\infty} Pr(x_n = j)$, which is denoted $\pi_j$, satisfies

$$\pi_j = \sum_{i=0}^{\infty} P_{j|i} \, \pi_i \quad , \quad j \geq 0, \tag{4.103}$$

and

$$\sum_{j=0}^{\infty} \pi_j = 1 \quad . \tag{4.104}$$

To relate the Markov chain $\{x_n\}$ to the calculation of $L^K_{\{T_k\}}(\{D_k\})$, a second Markov chain $\{y_n\}$ needs to be considered where

$$y_n = (x_{n+1}, x_n) \qquad n = 1, 2, \ldots \quad . \tag{4.105}$$

The transition probability of $\{y_n\}$ is given by

$$\begin{aligned} \hat{P}_{(j_1,j_2|i_1,i_2)} &= Pr(y_n = (j_1,j_2)|y_{n-1} = (i_1,i_2)) \\ &= Pr(x_n = j_1, x_{n-1} = j_2|x_{n-1} = i, x_{n-2} = i_2) \quad . \end{aligned} \tag{4.106}$$

Using (4.100), $\hat{P}_{(j_1,j_2|i_1,i_2)}$ is expressed in terms of $P_{j|i}$ by

$$\hat{P}_{(j_1,j_2|i_1,i_2)} = \begin{cases} P_{j_1|i_1} & \text{if} \quad j_2 = i_1 \\ 0 & \text{otherwise} \end{cases} \tag{4.107}$$

Then $\lim_{n\to\infty} Pr(y_n = (j_1,j_2))$, which is denoted $\hat{\pi}_{(j_1,j_2)}$, is given by

$$\hat{\pi}_{(j_1,j_2)} = P_{j_1|j_2}\, \pi_{j_2} \quad . \tag{4.108}$$

To compute $L^K_{\{T_k\}}(\{D_k\})$, functionals of the Markov chain $\{y_n\}$ are now defined. Let $N^k(y_n)$, $k = 1, \ldots, K$, be the number of messages which are compressed by compressor k during the time period $(t_n, t_{n+1}]$. It is easy to show that $N^k(y_n)$ as defined is a function of $x_n$ and $x_{n-1}$ and hence is a function of $y_n$. So, $N^k(y)$ is computed as

follows:

$$N^k(j_1,0) = \begin{cases} 1 & , \text{ if } \quad j_1 = \ell_1 \text{ and, } k = 1 \\ 0 & , \text{ otherwise} \end{cases} \quad , \tag{4.109}$$

$$N^k(j-1,j) = 0 \qquad \text{for all } j > 0 \text{ and for } k = 1,\ 2,\ \ldots,\ K \tag{4.110}$$

and for $j_2 > 0$, and $j_1 \geq j_2 - 1$

$$N^k(j_1 + m_k, j_2) = \begin{cases} N^k(j_1,j_2) + 1 & , \text{ if } \quad P_{j_1|j_2} \neq 0 \text{ and } m_{k-1} < j_2 \leq m_k \\ 0 & , \text{ if } \quad P_{j_1|j_2} = 0 \\ N^k(j_1,j_2) & , \text{ otherwise} \end{cases} \quad . \tag{4.111}$$

Then let $N(y_n)$ be the total number of arriving messages in the interval $(t_{n-1}, t_n]$, which is given by

$$N(y_n) = \sum_{k=1}^{K} N^k(y_n) \quad . \tag{4.112}$$

Now define the functional $W(y_n)$ to be the sum of the waiting times of messages arriving between $(t_{n-1}, t_n]$. So, $W(y_n)$ can be expressed in terms of $\{N^k(y_n)\}$ as follows:

$$(j_1,j_2) = \begin{cases} 0 & , \text{ if } \quad j_2 = 0 \\ 0 & , \text{ if } \quad j_2 > 0, \text{ and } j_1 = j_2 - 1 \\ N(j_1,j_2)(j_2 - .5)\Delta + \displaystyle\sum_{k=1}^{K} \frac{(N^k(j_1,j_2) - 1)\, N^k(j_1,j_2)}{2} m_k \Delta & \\ \quad + \displaystyle\sum_{k=2}^{N} \left[ \sum_{k_1=1}^{k} N^{k_1}(j_1,j_2)\, N^k(j_1,j_2) m_{k_1} \Delta \right] & , \text{ otherwise.} \end{cases} \tag{4.113}$$

Using the Markov chain $\{y_n\}$, the following theorem shows that

$$\bar{W} = \lim_{N\to\infty} N^{-1} E\{\sum_{n=0}^{N-1} (w_n)\} \tag{4.114}$$

and

$$P_k = \lim_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} \Pr(T_{k-1} < w_n \le T_k) \tag{4.115}$$

can be expressed in terms of the mean of the functionals of the Markov chain. Hence, $L^K_{\{T_k\}}(\{D_k\})$ can be evaluated from (3.15).

<u>Theorem 4.4</u>

Consider an adaptive data compression scheme, which uses K data compressors, where the greatest common divisor of the transmission times $\{t_k\}$ and thresholds $T_k$ is $\Delta$. Then $\bar{W}$ and $P_k$ are given by

$$\bar{W} = \frac{\sum_{j_1 j_2} W(j_1,j_2)\,\hat{\pi}_{j_1 j_2}}{\sum_{j_1 j_2} N(j_1,j_2)\,\hat{\pi}_{j_1 j_2}} \tag{4.116}$$

and

$$P_k = \frac{\sum_{j_1 j_2} N^k(j_1,j_2)\hat{\pi}_{j_1 j_2}}{\sum_{j_1 j_2} N(j_1,j_2)\hat{\pi}_{j_1 j_2}} \quad . \tag{4.117}$$

<u>Proof</u>

From the generalization of Lemma 3.10, it is clear that

$$\bar{W} = \lim_{n\to\infty} E(w_n) \tag{4.118}$$

and

$$P_k = \lim_{n\to\infty} \Pr(T_{k-1} < w_n \leq T_k) \quad , \quad k = 1, \ldots, K. \tag{4.119}$$

Using the ergodic theorem (Doob [38]), (4.118), and (4.119), it is clear that

$$\lim_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} w_n = \bar{W} \qquad \text{w.p.1} \tag{4.120}$$

and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} I_k(w_n) = P_k \qquad \text{w.p.1} \quad , \quad k = 1, \ldots, K \quad , \tag{4.121}$$

where $I_k(\cdot)$ is the indicator function defined by

$$I_k(w) = \begin{cases} 1 \ , & \text{if} \quad T_{k-1} < w \leq T_k \\ 0 \ , & \text{otherwise} \end{cases} \quad . \tag{4.122}$$

Now let $M_i$ be given by

$$M_i = \sum_{n=1}^{i} N(y_n) \quad . \tag{4.123}$$

So, from the definition of the sequence $\{\hat{t}_n\}$, (4.95), $M_i$ is lower bounded by considering the situation where every arrival initiates a busy period and so,

$$M_i \geq \frac{i}{\ell_1 + 1} \quad . \tag{4.124}$$

Thus,

$$\frac{\sum_{n=1}^{i} W(y_n)}{M_i} = \frac{\sum_{n=1}^{M_i} w_n}{M_i} \tag{4.125}$$

and

$$\frac{\sum_{n=1}^{i} N^k(y_n)}{M_i} = \frac{\sum_{n=1}^{M_i} I_k(w_n)}{M_i} \qquad k = 1, \ldots, N \quad . \tag{4.126}$$

From the bound (4.125), and from (4.120) and (4.121) it is clear that the limit of (4.125) and (4.126) are given by

$$\lim_{i\to\infty} \frac{\sum_{n=1} W(y_n)}{M_i} = \bar{W} \qquad \text{w.p.1} \tag{4.127}$$

and

$$\lim_{i\to\infty} \frac{\sum_{n=1}^{i} N^k(y_n)}{M_i} = P_k \qquad \text{w.p.1} \quad , \; k = 1, \ldots, K. \tag{4.128}$$

Now from an ergodic theorem in Chung [39],

$$\lim_{i\to\infty} \frac{\sum_{n=1}^{i} W(y_n)}{M_i} = \frac{\sum_{j_1 j_2} \hat{\pi}_{j_1 j_2} W(j_1,j_2)}{\sum_{j_1 j_2} \hat{\pi}_{j_1 j_2} N(j_1,j_2)} \qquad \text{w.p.1} \tag{4.129}$$

and

$$\lim_{i\to\infty} \sum_{n=1}^{i} \frac{N^k(y_n)}{M_i} = \frac{\sum_{j_1 j_2} \hat{\pi}_{j_1 j_2} N^k(j_1,j_2)}{\sum_{j_1 j_2} \hat{\pi}_{j_1 j_2} N(j_1,j_2)} \qquad \text{w.p.1}, \quad k = 1,2,\ldots,K \quad . \tag{4.130}$$

Hence, substituting (4.127) and (4.128) into (4.129) and (4.130), the expressions (4.116) and (4.117) are proved.

<u>Q.E.D</u>

Using (4.116) and (4.117), $\bar{W}$ and $\{P_k\}$ can be computed once

$\hat{\pi}_{j_1j_2}$ is computed or once $\hat{\pi}_{j_1j_2}/\pi_0$ is computed. From (4.110), $\hat{\pi}_{j_1j_2}/\pi_0$ is determined from $\pi_j/\pi_0$, $j = 0, 1, \ldots$ . Now clearly from (4.102)

$$P_{j|i} = 0$$

for i and j that satisfies $j + 1 \geq i \geq 0$. So, $\pi_j/\pi_0$ is given by solving (4.103) in a recursive manner. Thus,

$$\frac{\pi_{j+1}}{\pi_0} = \left(\frac{\pi_j}{\pi_0} - \sum_{i=0}^{j} P_{j|i}\frac{\pi_i}{\pi_0}\right)P_{j|j+1}^{-1} \quad , \quad j = 1, 2, \ldots \quad (4.131)$$

where $\dfrac{\pi_0}{\pi_0} = 1$.

Hence, $\overline{W}$ and $P_k$ can be computed using (4.116) and (4.117), respectively, and $L^K_{\{T_k\}}(\{D_k\})$ can be evaluated using (3.15). Figure 4.2 summarizes this computation. In Section 4.5, these techniques for the evaluation of $L^K_{\{T_k\}}(\{D_k\})$ are used to obtain numerical results for $\gamma^1_K(D)$, $K > 2$.

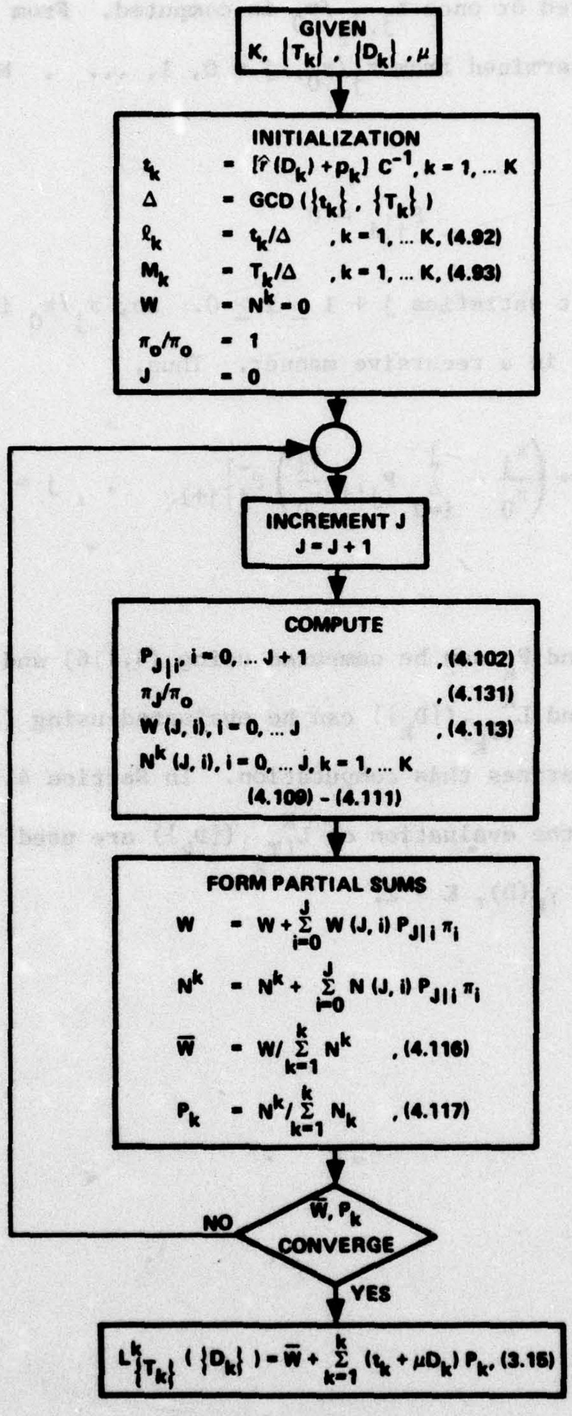



Figure 4.2. Evaluation of $L^{K}_{\{T_k\}}(\{D_k\})$, $K>2$

## 4.5 Numerical Results

To illustrate the performance improvement when using the adaptive data compression scheme as compared with the use of the nonadaptive scheme, a single channel communication system is considered. This comparison is performed by evaluating the delay distortion relationship for the different schemes. The source and channel model to be employed are described in Section 3.2. The message arrival rate is $\lambda[\frac{\text{mess.}}{\text{sec.}}]$ and the channel transmits information bits at a rate $C[\frac{\text{nats}}{\text{sec.}}]$. Each message consists of a collection of $\nu$ letters which are independent normally distributed random variables with variance $\sigma^2$. The data compressors used in the compression schemes under consideration satisfy the rate distortion relationhip $\hat{r}(D)$ $[\frac{\text{nats}}{\text{mess}}]$. Using a per-letter squared distortion criteria, $\hat{r}(D)$ is given by (see, Berger [4])

$$\hat{r}(D) = \nu/2 \ \ln(\sigma^2/D) \tag{4.132}$$

The protocol information parameter $p_K[\frac{\text{nats}}{\text{mess}}]$ is given by

$$p_K = \ln K \tag{4.133}$$

where K is the number of data compressors used in the compression scheme.

Using the models described, the delay distortion relationships $\gamma^1_K(D)$ for the nonadaptive compression scheme, K=1, and for the adaptive compression scheme, K=2, and 3, are evaluated using (3.10) and (4.7), respectively. In the evaluation of $\gamma^1_K(D)$, K = 2, 3, the quantity $L^2_{\{T_k\}}(\{D_k\})$ is computed using the technique described in

Figure 4.1 and the quantity $L^3_{\{T_k\}}(\{D_k\})$ is computed using the technique described in Figure 4.2. The optimization required in the evaluation of $\gamma^1_K(D)$ is performed numerically using a numerical optimization algorithm. Figure 4.3 presents curves for $\gamma^1_K(D)$, K = 1, 2, 3. Figure 4.4 shows the relative performance of the nonadaptive scheme versus the adaptive scheme by presenting the curves $\gamma^1_1(D) - \gamma^1_K(D)$, K = 2, 3.

From the figures it is evident that at low distortion levels the adaptive schemes yield considerable smaller message delays than those obtained when employing nonadaptive schemes. At high distortion levels, nonadaptive schemes are slightly superior to the adaptive schemes, due to the required transmission of protocol information for the adaptive data compression schemes. The lower envelope of the curves in Figure 4.3 represent a restricted version of $\gamma_A(D)$ defined in (3.14) where only schemes with one, two, or three data compressors are considered in the calculation of $\gamma_A(D)$.

Thus, from this example, it is concluded that the adaptive data compression schemes considered here serve effectively as a device to reduce message delays when low distortion levels are required. The situation of low distortion levels corresponds to heavy traffic in the communication network. As the distortion levels increase, message delays associated with the corresponding adaptive compression schemes are slightly degraded when compared with those obtained by nonadaptive schemes. This is due to the increased traffic resulting from the transmission of protocol information when using the adaptive scheme.

Figure 4.1 and the quantity $L^3_{\{T_k\}}(\{D_k\})$ is computed using the technique described in Figure 4.2. The optimization required in the evaluation of $\gamma^1_K(D)$ is performed numerically using a numerical optimization algorithm. Figure 4.3 presents curves for $\gamma^1_K(D)$, K = 1, 2, 3. Figure 4.4 shows the relative performance of the nonadaptive scheme versus the adaptive scheme by presenting the curves $\gamma^1_1(D) - \gamma^1_K(D)$, K = 2, 3.

From the figures it is evident that at low distortion levels the adaptive schemes yield considerable smaller message delays than those obtained when employing nonadaptive schemes. At high distortion levels, nonadaptive schemes are slightly superior to the adaptive schemes, due to the required transmission of protocol information for the adaptive data compression schemes. The lower envelope of the curves in Figure 4.3 represent a restricted version of $\gamma_A(D)$ defined in (3.14) where only schemes with one, two, or three data compressors are considered in the calculation of $\gamma_A(D)$.

Thus, from this example, it is concluded that the adaptive data compression schemes considered here serve effectively as a device to reduce message delays when low distortion levels are required. The situation of low distortion levels corresponds to heavy traffic in the communication network. As the distortion levels increase, message delays associated with the corresponding adaptive compression schemes are slightly degraded when compared with those obtained by nonadaptive schemes. This is due to the increased traffic resulting from the transmission of protocol information when using the adaptive scheme.

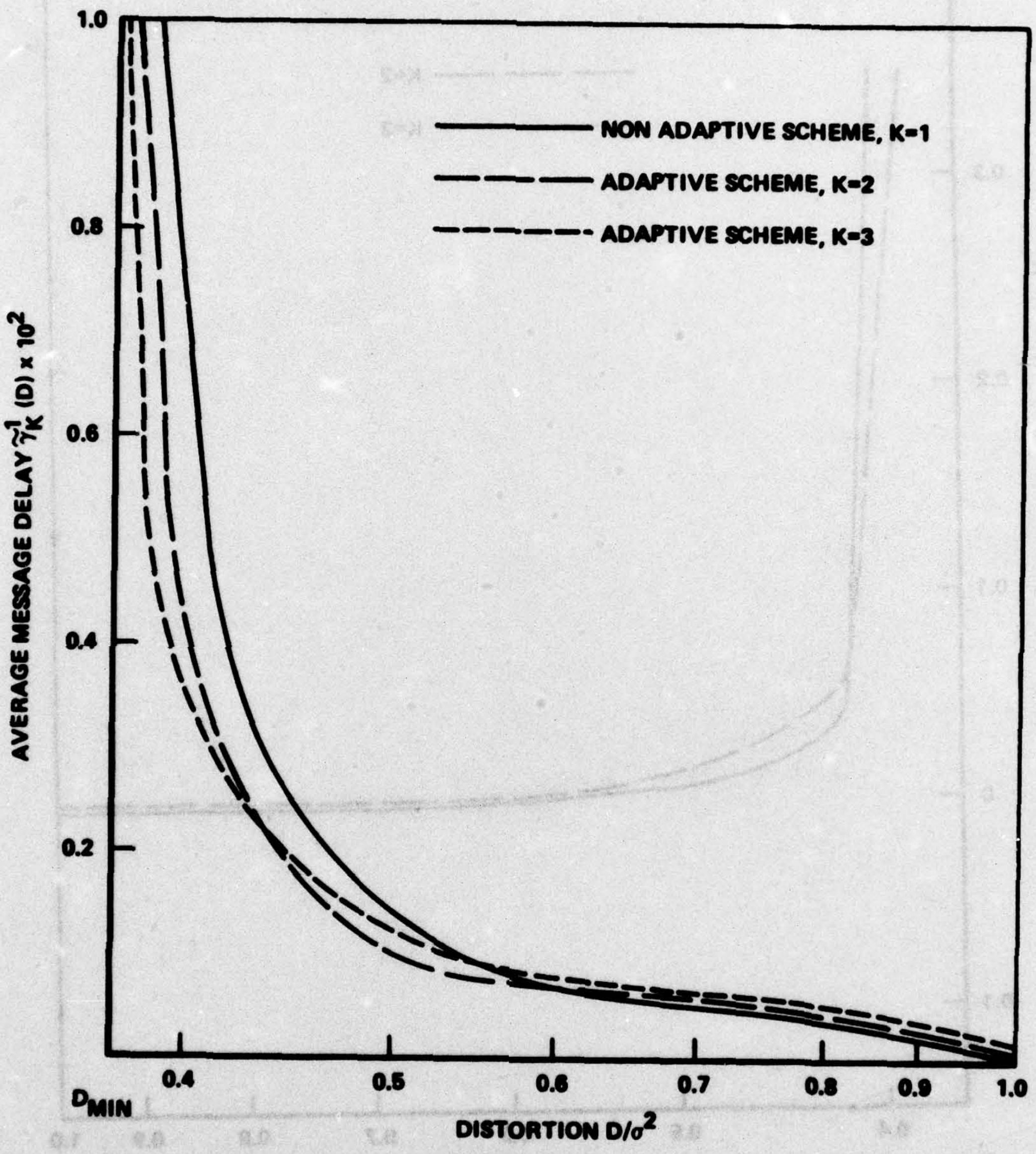



**Figure 4.3. Delay Distortion Curves, $\tilde{\gamma}_K^{-1}$ (D), for a Single Channel System, Gaussian Source, $\nu$ = 100 [Letter/Mess.], C = [Knats/Sec.], $\lambda$ = 1000 [Mess./Sec.], with Parameter K the Number of Data Compressors Employed**

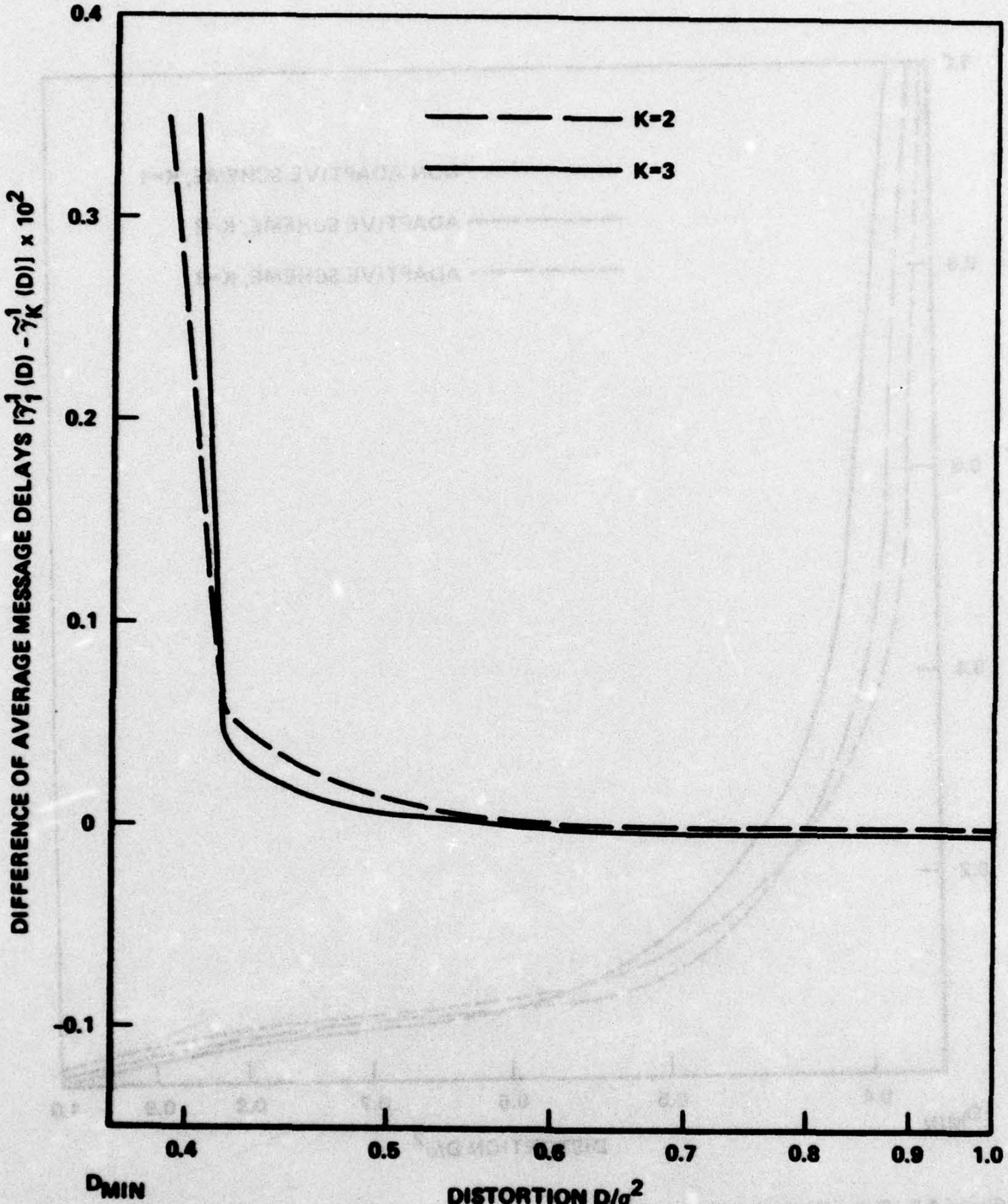



**Figure 4.4.** Difference of Delay Distortion Curves, $\tilde{\gamma}_1^1(D) - \tilde{\gamma}_K^1(D)$, for a Single Channel System, Gaussian Source, $\nu$ = 100 [Letters/Mess.], C = 50 [Knats/Sec.], $\lambda$ = 1000 [Mess./Sec.], with Parameter K the Number of Data Compressors Employed

### 4.6 Conclusions

The delay distortion relationship for a communication system utilizing a single channel and an adaptive data compression scheme was considered in this chapter. The delay distortion relationship was shown to be specified by a minimization of a functional over a vector space. Basic properties of this relationship were presented. The functional to be minimized was then examined, and an integral equation whose solution is related to this functional was derived. For the case of an adaptive data compression scheme which uses two data compressors, this integral equation was solved and the functional evaluated. For the case of compression schemes which use more than two data compressors, a technique was developed to evaluate the functional by defining an appropriate Markov chain and functionals of the Markov chain. Finally, examples of the delay distortion relationship for a communication network employing a single communication channel were presented. From the examples, it was observed that an adaptive compression scheme yields superior performance to that of a nonadaptive scheme at low distortion levels.

# CHAPTER V

# THE DELAY DISTORTION RELATIONSHIP FOR TANDEM CHANNELS

The results in Chapters III and IV pertain to a communication network which uses a single communication channel for data transmission. In this chapter, these results are extended to communication networks employing tandem channels where the channel capacities are all equal. In Section 5.1, the tandem channel network is discussed and the delay distortion relationship for this network is determined. The resulting delay distortion relationship is shown to be of the same form as the delay distortion relationship for the network consisting of a single channel. In Section 5.2, some of the properties of the delay distortion relationship outlined in Section 5.1 are presented. Finally in Section 5.3, numerical results for the delay distortion relationship for various tandem networks are presented.

## 5.1 Delay Distortion Relationship for Tandem Channels

In this section, the delay distortion relationship is investigated for a network configuration consisting of a tandem channel, as described in Section 3.1; a source, as described in Section 3.1; and an adaptive data compression scheme, as described in Section 3.2. Incorporating the results of Chapters III and IV, this relationship is shown to have a form similar to the delay distortion relationship for the single channel network. Thus, the results of this section

will closely follow the results of the previous chapters.

The network under consideration consists of M channels, with identical channel capacities equal to C [nats/sec], connected in a tandem manner as depicted in Figure 3.1. For this network, and for the adaptive data compression scheme which uses two data compressors, the delay distortion relationship, $\gamma_2^M(D)$, is given by (3.13). As in Section 3.2, if the infimum in (3.13) can be replaced with the minimum operation, the expression for $\gamma_2^M(D)$ reduces to (3.16). The term $L_\pi^2(\{D_k\})$ in (3.16) is expressed using (3.11), (3.12) and (3.15) as

$$L_\pi^2(\{D_k\}) = \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_\pi\{ \sum_{n=1}^{n} [w_{n,m} + (\hat{r}(d_n) + p_2)C^{-1}] + \mu d_n\} \tag{5.1}$$

where $\pi$ is the compression assignment policy, $\{D_k\}$ is the average distortion levels of the data compressors used, $w_{n,m}$ is the waiting time of the $n^{th}$ message in the buffer associated with the $m^{th}$ channel with $w_{0,m} = 0$ for m = 1, 2, ..., M (network initially empty), $d_n$ is the average distortion associated with the data compressor which compresses the $n^{th}$ message, $\mu$ is the associated Lagrange multiplier, $\hat{r}(D)$ is a rate distortion relationship that the data compressors satisfy and $p_2$ is the protocol information. Without loss of generality, assume $D_1 \le D_2$, for the remainder of this chapter.

To determine if (3.16) is valid, it is necessary to investigate

$$L^2(\{D_k\}) = \inf_{\pi\in\Pi} L_\pi(\{D_k\}) \tag{5.2}$$

and ascertain if there exists a $\pi^* \in \Pi$ which satisfies

$$L^2(\{D_k\}) = L^2_{\pi^*}(\{D_k\}) \qquad (5.3)$$

where $\Pi$ is the class of causal policies. As in the case of the single communication channel in Chapter III, the verification of (5.3) is accomplished by introducing the notion of a Markov decision process for the tandem communication channel, and then determine $\pi^*$ as in Section 3.4. Once (3.16) is verified, optimization techniques are used to find $\gamma_2^M(D)$.

The Markov decision process $\{((x_n, a_n), n = 0, 1, 2, \ldots\}$ which is considered has a state $x_n$ described by three elements. The first element, to be henceforth denoted as $w_n$, is the waiting time, $w_{n,1}$, of the $n^{th}$ message in the buffer associated with the first channel. The second element, denoted $i_n$, consists of the maximum of $I_1$ and of the sum of the idle periods of channel one since the last transmission of a message compressed by data compressor one where

$$I_1 = (M-1)\,(t_1 - t_2) \qquad (5.4)$$

and $t_k$ is the transmission time of a message compressed by data compressor k which is given by

$$t_k = (\hat{r}(D_k) + p_2)C^{-1} \quad . \qquad (5.5)$$

The last element of $x_n$, denoted $\tau_n$, consists of the interarrival time between the $(n-1)^{st}$ message and the $n^{th}$ message. So, $x_n = (w_n, i_n, \tau_n)$, and the state space is $X = ([0,\infty) \times [0,I_1] \times [0,\infty))$.

The action at the $n^{th}$ decision epoch is $a_n \in [(1),(2)]$ which

corresponds to using data compressor 1 or 2 on the $n^{th}$ message. Then the history, denoted $H_n$, of the Markov decision process up to the arrival of the $n^{th}$ message, is given by the sequence of states and actions as

$$H_n = (x_0, a_0, x_1, a_1, \ldots x_n, a_n) \quad . \tag{5.6}$$

The history pattern $H_n$ may not be the complete history of the communication system. A sufficient condition for $H_n$ to be the history of the communication system is that the state $x_n$ specifies the waiting times $w_{n,m}$, $m = 1, \ldots, M$ of the $n^{th}$ message at each of the M channels. The following lemma proves this.

Lemma 5.1

Let the initial waiting times at the various nodes satisfy

$$w_{0,1} = w_0 \tag{5.7}$$

and for $m = 2, \ldots, M$ satisfy

$$w_{0,m} = 0 \tag{5.8}$$

or

$$w_{0,m} = \max\{0, t_1 - t_2 - \max(0, i_0 - (m-2)(t_1 - t_2))\} \tag{5.9}$$

where the initial state of the Markov decision process is $x_0 = (w_0, i_0, \tau_0)$. Then the waiting times at the various nodes are given for $n > 0$ by

$$w_{n,1} = w_n \tag{5.10}$$

and for m = 2, ..., M by

$$w_{n,m} = \begin{cases} 0 & ,\text{if } a_n = 1 \\ \max[0,\ t_1 - t_2 - \max(0,\ i_n - (m-2)(t_1 - t_2))] & ,\text{if } a_n = 2 \end{cases} \tag{5.11}$$

where

$$i_n = \begin{cases} \min[I_1,\ \max(0,\ \tau_n - w_{n-1} - t_1)] & ,\ \text{if } a_n = 1 \\ \min[I_1,\ i_n + \max(0,\ \tau_n - w_{n-1} - t_2)] & ,\ \text{if } a_n = 2 \end{cases} \tag{5.12}$$

and $x_n = (w_n, i_n, \tau_n)$ is the state of the Markov decision process.

**Proof**

The proof of the lemma is by induction. Suppose (5.10) and (5.11) are valid for the $n^{th}$ message. By the recurrance relationship for waiting times, $w_{n+1,m}$ is given by

$$w_{n+1,m} = \max(0,\ w_{n,m} + t_{a_n} - \tau^m_{n+1}) \tag{5.13}$$

where $\tau^m_n$ is the interarrival time between the arrival of the $(n-1)^{st}$ message and the $n^{th}$ message to the $m^{th}$ channel.

First consider the case $a_{n+1} = 1$. Clearly, from (5.12) for $m \geq 2$

$$w_{n,m} + t_{a_n} \leq t_1 \tag{5.14}$$

independent of the value of $a_n$. Furthermore,

$$\tau^m_{n+1} \geq t_1 \qquad (5.15)$$

since the transmission time of a message which is compressed by data compressor one through a single channel is $t_1$. So, using (5.14) and (5.15) in (5.13)

$$w_{n+1,m} = 0 \qquad (5.16)$$

for $m \geq 2$.

For the second case of $a_{n+1} = 2$, let $m \geq 2$. The interarrival time $\tau^m_n$ *for all* $n \geq 0$ *is given by*

$$\tau^m_n = I^{m-1}_n + t_{a_n} \qquad (5.17)$$

where $I^m_n$ is the idle time of $m^{th}$ channel between transmissions of the $(n-1)^{st}$ message and the $n^{th}$ message. But $I^m_{n+1}$ is given by

$$I^m_{n+1} = \max(0, \tau^m_{n+1} - t_{a_n} - w_{n,m}) \quad . \qquad (5.18)$$

Upon substituting (5.17) into (5.18), $I^m_{n+1}$ is given by the recurrence relationship

$$I^m_{n+1} = \max(0, I^{m-1}_{n+1} - w_{n,m} + t_2 - t_{a_n}) \quad . \qquad (5.19)$$

So, from (5.17), the expression for $\tau^m_{n+1}$ becomes

$$\tau^m_{n+1} = \max(0, I^{m-1}_{n+1} - w_{n,m} - t_{a_n} + t_2) + t_2 \quad . \qquad (5.20)$$

Substituting (5.20) into (5.13) results in

$$w_{m+1,m} = \max\{ 0, w_{n,m} + t_{a_n} - t_2 - \max[0, I^{m-1}_{n+1} - w_{n,m} - t_{a_n} + t_2]\} \quad . \qquad (5.21)$$

Now for $a_n = 1$, substituting (5.11) in (5.21), $w_{n+1,m}$ is given by

$$w_{n+1,m} = \max\{0,\ t_1 - t_2 - \max[0,\ I^{m-1}_{n+1} - (t_1 - t_2)]\} \quad . \tag{5.22}$$

But from (5.19) for $a_n = 1$

$$I^m_{n+1} = \max[0,\ I^1_{n+11} - (m-1)\ (t_1 - t_2)] \tag{5.23}$$

and so, upon substitution into (5.22),

$$w_{n+1,m} = \max\{0,\ t_1 - t_2 - \max[I^1_{n+1} - (m-2)\ (t_1 - t_2)]\} \tag{5.24}$$

For $a_n = 2$, substituting (5.11) and (5.17) into (5.25), $w_{n+1,m}$ is given by

$$w_{n+1,m} = \max\{0,\ t_1 - t_2 - \max[0,\ i_n - (m-2)\ (t_1 - t_2)] - I^{m-1}_{n+1}\} \tag{5.25}$$

where $I^m_{n+1}$ is given by the solution of (5.19) as

$$I^m_{n+1} = \max(0,\ I^1_{n+1} - \sum_{m'=2}^{m-1} w_{n,m'}) \quad . \tag{5.26}$$

Now consider three cases. First let m be such that $w_{n,m} = 0$. Hence, $w_{n,m'} = 0$ for $m \geq m' \geq 2$ and so

$$w_{n+1,m} = \max\{0,\ t_1 - t_2 - \max[0,\ i_n + I^1_{n+1} - (m-2)\ (t_1 - t_2)]\} \ . \tag{5.27}$$

Then let $\hat{m} \geq 2$ be the smallest integer such that $w_{n,\hat{m}} > 0$. So, $i_n - (\hat{m}-2)\ (t_1 - t_2) \geq 0$, thus, $w_{n+1,\hat{m}}$ satisfies (5.27) since $I^{\hat{m}-1}_{n+1} = I^1_{n+1}$. Finally let $m > \hat{m}$. Since for all $m' > \hat{m}$, $w_{n,m'} = t_1 - t_2$, $I^m_{n+1}$ is given by

$$I^{m}_{n+1} = \max(0,\ I^{1}_{n+1} - w_{n,\hat{m}} \sum_{m'=\hat{m}+1}^{m-1} (t_1 - t_2)) \quad . \qquad (5.28)$$

Substituting the values of $w_{n,\hat{m}}$ into (5.28), $I^{m}_{n+1}$ is given by

$$I^{m}_{n+1} = \max\{0,\ I^{1}_{n+1} - [t_1 - t_2 - i_n + (\hat{m}-2)\ (t_1 - t_2)] - \sum_{m'=\hat{m}+1}^{m-1} (t_1 - t_2)\}$$

$$= \max\{0,\ i_n + I^{1}_{n+1} - (m-2)\ (t_1 - t_2)\} \quad . \qquad (5.29)$$

Hence substituting (5.29) into (5.25) and noting that $i_n - (m-2) \cdot (t_1 - t_2) \leq 0$, expression (5.27) results. Therefore for all $m \geq 2$, expression (5.27) is valid. Now observe that in (5.27) $i_n + I^{1}_{n+1}$ can be replaced by $\max(I_1,\ i_n + I^{1}_{n+1})$ without affecting the value of $w_{n+1,m}$. Clearly, the element $i_{n+1}$ of $x_{n+1}$ is given by

$$i_{n+1} = \begin{cases} \max(I_1,\ I^{1}_{n+1}) & \text{if} \quad a_n = 1 \\ \max(I_1,\ i_n + I^{1}_{n+1}) & \text{if} \quad a_n = 2 \ . \end{cases} \qquad (5.30)$$

Hence, (5.11) is valid for n+1. Therefore, by induction since $w_{0,m}$ specified by (5.7) - (5.9) satisfies (5.11), then (5.11) is valid for all n.

Q.E.D

Lemma 5.1 shows that history $H_n$ is the complete history of the communication system for all initial conditions of the form (5.8) or (5.9). Certaintly, the initial condition requiring all the buffers in the communication system to be empty, which is assumed in calculating $\gamma^{M}_{2}(D)$, satisfies (5.8) when $i_0 = I_1$. Thus, initial conditions other than that of the form of 5.8) or (5.9) need not be considered.

Using the recursive equation (3.21) for $w_n$, and the recursive equation (5.12) for $i_n$, a version of the transition probability distribution, $Pr(w_{n+1} \le W,\ i_{n+1} \le I,\ \tau_{n+1} \le T | H_n)$ is given by

$$Pr(w_{n+1} \le W,\ i_{n+1} \le I,\ \tau_{n+1} \le T | H_n)$$

$$= Pr(w_{n+1} \le W,\ i_{n+1} \le I,\ \tau_{n+1} \le T | (w_n, i_n, \tau_n), a_n)$$

$$= \begin{cases} F_\tau(\min(I+w_n+t_1, T)) - F_\tau(\max(0\ , w_n+t_{a_1}-W)) & , \text{if } a_n=1,\ I \le I_1 \\ F_\tau(T) - F_\tau(\max(0\ , w_n+t_{a_1}-W)) & , \text{if } a_n=1,\ I > I_1 \\ F_\tau(\min(I-i_n+w_n+t_2, T)) - F_\tau(\max(0\ , w_n+t_2-W)), & \text{if } a_n=2,\ I \le I_1 \\ F_\tau(T) - F_\tau(\max(0\ , w_n+t_2-W)) & , \text{if } a_n=2,\ I > I_1\ . \end{cases} \tag{5.31}$$

The rules or policies $\pi$, associated with the Markov decision process, are specified in Section 3.3, and the set $\Pi$ is a collection of such policies. The cost function $c(x,a)$ is of the form

$$c(x,a) = \begin{cases} w + Mt_1 + \mu D_1 & , \text{ if } \quad a = 1 \\ w + t_2 + (M-1)t_1 - i + \mu D_2 & , \text{ if } \quad a = 2 \end{cases} \tag{5.32}$$

where $x = (w,i,\tau) \in X$, and $a \in [(1),(2)]$. Hence the average cost $\Phi(x,\pi)$ under policy $\pi \in \Pi$ when the Markov decision process $(x_n, a_n)$ is initially in state $x \in X$ is given by (3.39) to be

$$\Phi(x,\pi) = \limsup_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} E_{\hat{\pi}}\{c(x_n, a_n) | x_0 = x\} \tag{5.33}$$

and

$$\Phi(x) = \inf_{\pi\varepsilon\Pi} \Phi(x,\pi) \quad . \tag{5.34}$$

Relating this to the communication system under consideration, it is clear from Lemma 3.1 and (5.32) that

$$\sum_{n=1}^{M} w_{n,m} + Mt_{a_n} + \mu D_{a_n} = c(x_n, a_n) \quad . \tag{5.35}$$

Thus, by properly initializing the Markov decision process $(x_n, a_n)$, $L_\pi^2(\{D_k\})$ is clearly given by

$$L_\pi^2(\{D_k\}) = \Phi((0, I_1, 0), \pi) \quad . \tag{5.36}$$

Since $H_n$ is the complete history of the communication system, the set of admissible policies $\Pi$ is the same set of policies considered in (5.2). Hence, $L^2(\{D_k\})$ is given by

$$L^2(\{D_k\}) = \Phi((0, I, 0)) \quad . \tag{5.37}$$

Therefore, the problem of finding a policy $\pi^*$ which satisfies (5.3) reduces to the problem of locating a policy $\pi^*$ which satisfies

$$\Phi(x) = \Phi(x,\pi^*) \quad . \tag{5.38}$$

Before determining a $\pi^*$ which satisfies (5.38), the value of $D_2$ must be restricted to guarantee that $\Phi(x)$ is finite. So, as in the single channel network case in Section 3.4, $D_2$ is restricted to the region

$$D_2 < \hat{r}^{-1}(\frac{c}{\lambda} - p_2) \quad .$$

Outside this region $\Phi(x)$ is unbounded.

To find $\pi^*$ and $\Phi(x)$, first consider the Markov decision processes $\{(x_n, a_n), n = 0, 1, 2, \ldots\}$ defined as before, but with cost function $c(x,a)$, given by (5.32), replaced by

$$c'(x,a) = \begin{cases} w + Mt_1 + \mu D_1 & ,\text{if} \quad a = 1 \\ w + t_2 + (M-1)t_1 + \mu D_2 & ,\text{if} \quad a = 2 \text{ and } i = 0 \\ w + Mt_2 + \mu D_2 & ,\text{if} \quad a = 2 \text{ and } i > 0 \end{cases} \tag{5.39}$$

where $x = (w, i, \tau)$. The associated average cost is denoted $\Phi'(x,\pi)$ and $\Phi'(x)$ is given by

$$\Phi'(x) = \inf_{\pi\varepsilon\Pi} \Phi'(x,\Pi) \quad . \tag{5.40}$$

Since from (5.32) and (5.39) for all $x \varepsilon X$ and $a \varepsilon [(1),(2)]$

$$c'(x,a) \leq c(x,a) \quad , \tag{5.41}$$

then

$$\Phi'(x,\pi) \leq \Phi(x,\pi) \tag{5.42}$$

for all $\pi \varepsilon \Pi$. Thus, it is clear that for all $x \varepsilon X$

$$\Phi'(x) \leq \Phi(x) \quad . \tag{5.43}$$

So, from (5.43), if a policy $\pi' \varepsilon \Pi$ can be found such that $\Phi(x,\pi')$ equals $\Phi'(x)$ for all $x \varepsilon X$, then $\pi'$ is the policy $\pi^*$ which satisfies (5.36). A logical choice for $\pi'$ is for $\pi'$ to satisfy

$$\Phi'(x) = \Phi'(x,\Pi') \quad , \tag{5.44}$$

if such a policy exists. In order to find $\pi' \varepsilon \Pi$ that satisfies (5.44), Theorem 3.2 is to be used. The following set of lemmas verify Assumptions 3.1, 3.2 and 3.3 which are required to apply Theorem 3.2.

<u>Lemma 5.2</u>

Assumption 3.1 is valid for the Markov decision process $(x_n, a_n)$ under consideration with costs $c'(x,a)$ given by (5.39).

<u>Proof</u>

Let m = 1,

$$g((w, i, \tau)) = w + 1 \tag{5.45}$$

and

$$b = \max_{k \varepsilon [1,2]} t_k \quad . \tag{5.46}$$

Then (3.42) and (3.43) are easily verified. Hence the assumption is valid.

<u>Q.E.D</u>

From Theorem 3.1, there exists α-optimal stationary deterministic policies for α ε (0,1) which can be obtained using the policy improvement algorithm of Corollary 3.1. The following lemma establishes the structure of the α-optimal policy.

**Lemma 5.3**

The α-optimal policy for the Markov decision process $(x_n, a_n)$ under consideration with cost function $c'(x,a)$ given by (5.39) is a connected policy where action one is selected when the waiting time, w, and the truncated sum of the idle times, i, satisfy

$$w \leq T_\alpha(i) \qquad (5.47)$$

and action two is selected otherwise. $T_\alpha(i)$ is termed the threshold function, and is selected to minimize the α-discounted cost problem. Furthermore, $T_\alpha(i)$ is a constant for all $i \in (0,I_1]$ and $T_\alpha(0) \geq T_\alpha(I_1)$.

**Proof**

First, as in Lemma 3.4, one can show that the dependences on the interarrival time can be dropped. So, using the value of $c'((w,i,\tau),k)$, (5.39) and the expression for $Pr(w_{n+1} \leq W,\ i_{n+1} \leq I,\ \tau_{n+1} \leq T \mid (w_n,i_n,\tau_n),k)$, (5.31), the policy improvement algorithm (3.45) becomes for $(w, i, \tau) \in X$

$$U_{n+1,\alpha}(w,i) = \min_{a\in[(1),(2)]} \Big\{ w + g(i,a) + \alpha \int_0^{w+t_a} U_{n,\alpha}(w + t_a - \hat{w},\ i\delta_a)\, dF_\tau(\hat{w}) + \alpha \int_{w+t_a}^{\infty} U_{n,\alpha}(0,\ \min(I_1,\ i\delta_a + \hat{w} - w - t_a)\, dF_\tau(\hat{w}) \Big\} \qquad (5.48)$$

where

$$U_{0,\alpha}(w,i) = 0 \quad , \tag{5.49}$$

$$g(i,a) = \begin{cases} t_a + \mu D_a + (M-1)t_1 & , \quad \text{if } i = 0 \\ Mt_a + \mu D_a & , \quad \text{if } i > 0 \end{cases} \tag{5.50}$$

and

$$\delta_a = \begin{cases} 0 & , \quad \text{if} \quad a = 1 \\ 1 & , \quad \text{if} \quad a = 2 \end{cases} \quad . \tag{5.51}$$

Now define $\hat{U}_{n,\alpha}(w,i)$ by

$$\hat{U}_{n,\alpha}(w,i) = \int_0^w U_{n,\alpha}(w - \hat{w},i)\, dF_\tau(\hat{w}) + \int_w^\infty U_{n,\alpha}(0,\min(I_1,i+\hat{w}-w))\, dF_\tau(\hat{w}) \; . \tag{5.52}$$

Thus, (5.48) becomes

$$U_{n+1,\alpha}(w,i) = \min_{a\varepsilon[(1),(2)]} \{w + g(i,a) + \alpha\, \hat{U}_{n,\alpha}(w + t_a,\, i\delta_a)\} \tag{5.53}$$

To establish the limiting form of $U_{n,\alpha}(w,i)$ as $n \to \infty$, it is required to show that for all n and for all $\alpha \; \varepsilon \; (0,1)$, $U_{n,\alpha}(w,i)$ is a nondecreasing function of w for fixed i, and is a nonincreasing function of i for fixed w. First define $F$ to be the set of functions of two variables of the form $f(w,i)$ where $(w,i) \; \varepsilon \; ([0,\infty) \times [0,I_1])$ such that $f(w,i)$ is a nondecreasing function of w for i fixed, and $f(w,i)$ is a nonincreasing function of i for w fixed. Clearly, $U_{0,\alpha}(\cdot,\cdot) \; \varepsilon \; F$. Now suppose $U_{n,\alpha}(\cdot,\cdot) \; \varepsilon \; F$ and then from (5.52)

$\hat{U}_{n,\alpha}(\cdot,\cdot) \in F$. Thus, from (5.53) $U_{n+1,\alpha}(\cdot,\cdot) \in F$ for all $n \geq 0$ and for all $\alpha \in (0,1)$.

Now to show that the $\alpha$-optimal policy is a connected policy, define for $w < 0$

$$U_{n,\alpha}(w,i) = U_{n,\alpha}(0, \min(I_1, i - w)) \tag{5.54}$$

and

$$\hat{U}_{n,\alpha}(w,i) = \hat{U}_{n,\alpha}(0, \min(I_1, i - w) \quad . \tag{5.55}$$

Furthermore, define for $(w,i) \in ((-\infty,\infty) \times [0,I_1])$

$$\Delta U_{n,\alpha}(w,i) = U_{n,\alpha}(w + t_1,0) - U_{n,\alpha}(w + t_2,i) \tag{5.56}$$

and

$$\Delta\hat{U}_{n,\alpha}(w,i) = \hat{U}_{n,\alpha}(w + t_1,0) - \hat{U}_{n,\alpha}(w + t_2,i) \quad . \tag{5.57}$$

Clearly from (5.52), (5.56) and (5.57), $\Delta\hat{U}_{n,\alpha}(w,i)$ is given by

$$\Delta\hat{U}_{n,\alpha}(w,i) = \int_0^\infty \Delta U_{n,\alpha}(w - \hat{w},i)\, F_\tau(d\hat{w}) \quad . \tag{5.58}$$

Thus, from (5.53), for the (n+1) stage problem, the policy at the first stage is to select data compressor 1 if $w \geq 0$ and $i \in [0,I_1]$ satisfies

$$g(i,1) - g(i,2) + \alpha\, \Delta\hat{U}_{n,\alpha}(w,i) \leq 0 \tag{5.59}$$

and use data compressor 2 otherwise. So, the optimal policy at the first stage is a connected policy if $\Delta\hat{U}_{n,\alpha}(w,i) \in F'$ where $F'$ is the

set of functions $f(w,i)$ defined on $((-\infty,\infty) \times [0,I_1])$ which are non-decreasing functions of $w$ and $i$. Now define $F''$ to be a subset of $F'$ such that if $f(\cdot,\cdot) \in F''$ then for a fixed value of $w \in (-\infty,\infty)$ $f(w,i)$ is a constant for $i \in (0,I_1]$. Hence, if $\Delta\hat{U}_{n,\alpha}(\cdot,\cdot) \in F''$, (5.59) implies that the optimal policy is a connected policy. To show that $\Delta U_{n,\alpha}(\cdot,\cdot) \in F''$ for all $n \geq 0$ and all $\alpha \in (0,1)$, first observe that $\Delta U_{0,\alpha}(\cdot,\cdot) \in F''$ and, so by (5.58) $\Delta\hat{U}_{0,\alpha}(\cdot,\cdot) \in F''$. Now suppose $\Delta U_{n,\alpha}(\cdot,\cdot) \in F''$, and hence $\Delta\hat{U}_{n,\alpha}(\cdot,\cdot) \in F''$. This implies that at the first stage of the (n+1) stage problem, the optimal policy for a discount factor of $\alpha \in (0,1)$ is a connected policy with threshold function $T_{n+1,\alpha}(i)$ applied to the waiting time w. From (5.59) it is clear that $T_{n+1,\alpha}(i)$ is a constant for $i \in (0,I_1]$ and $T_{n+1,\alpha}(0) \geq T_{n+1,\alpha}(I_1)$. Using (5.53) in (5.56), $\Delta U_{n+1,\alpha}(w,i)$ for $(w,i) \in ((-\infty,\infty) \times [0,I_1])$ is given by

$$\Delta U_{n+1,\alpha}(w,i) = \begin{cases} \alpha U_{n+1,\alpha}(w+t_1,i) - \alpha U_{n+1,\alpha}(0,I_1) & \text{, if } w < -t_1 \\[2ex] t_1-t_2 + \alpha\Delta\hat{U}_{n,\alpha}(w+t_1,0) & \text{, if } w+t_1 \leq T(0),\ 0 < w+t_2 \leq T(i) \\[2ex] 2(t_1-t_2) + \mu(D_1-D_2) + I_1u(-i) + \alpha\Delta\hat{U}_{n,\alpha}(w+t_1,0) & \\ \qquad + \Delta\hat{U}_{n,\alpha}(w+t_2,i) & \text{, if } w+t_1 \leq T(0),\ w+t_2 \geq T(i) \\[2ex] \mu(D_2-D_1) & \text{, if } w+t_1 > T(0),\ 0 < w+t_2 \leq T(i) \\[2ex] (t_1-t_2) + I_1u(-i) + \alpha\Delta\hat{U}_{n,\alpha}(w+t_2,i) & \text{, if } w+t_1 > T(0),\ w+t_2 > T(i) \end{cases} \tag{5.60}$$

where $u(\cdot)$ is the unit step function defined by

$$u(i) = \begin{cases} 0 & , \text{ if } \quad i < 0 \\ 1 & , \text{ if } \quad i \geq 0 \end{cases} \quad . \tag{5.61}$$

Since $\hat{U}_{n,\alpha}(w,0)$ is an increasing function of $w$, $\Delta\hat{U}_{n,\alpha}(\cdot,\cdot) \in F''$, and $U_{n+1,\alpha}(w,i)$ is a continuous function of $w$, then it is clear from (5.60) that $\Delta U_{n+1,\alpha}(\cdot,\cdot) \in F''$. Hence by induction, $\Delta U_{n,\alpha}(\cdot,\cdot) \in F''$ for all $n \geq 0$.

Now from (3.46) the $\alpha$-discounted cost using the $\alpha$-optimal policy $\pi^*_\alpha$ is given for $(w,i) \in ([0,\infty) \times [0,I_1])$ by

$$V_\alpha((w,i),\pi^*_\alpha) = \lim_{n\to\infty} U_{n,\alpha}(w,i) \tag{5.62}$$

where the dependences on the interarrival time $\tau$ have been dropped. Then $\Delta V_\alpha((w,i),\pi^*_\alpha)$, defined by

$$\begin{aligned} \Delta V_\alpha((w,i),\pi^*_\alpha) &= V_\alpha((w + t_1,0),\pi^*_\alpha) - V_\alpha((w + t_2,i),\pi^*_\alpha) \\ &= \lim_{n\to\infty} \Delta U_{n,\alpha}(w,i) \quad , \end{aligned} \tag{5.63}$$

is a limit of functions which are elements of $F''$. Hence, $\Delta V_\alpha((\cdot,\cdot),\pi^*_\alpha) \in F''$. Furthermore, from (3.44) and (5.39) $\pi^*_\alpha$ is the policy which selects data compressor 1 if $(w,i)$ satisfies

$$t_1 - t_2 + \mu(D_1 - D_2) + I_1 u(-i) + \alpha \int_0^\infty \Delta V_\alpha(w - \hat{w},i) F_\tau(d\hat{w}) \leq 0 \tag{5.64}$$

and selects data compressor 2 otherwise. Since $\Delta V_\alpha((\cdot,\cdot),\pi^*_\alpha) \in F''$, expression (5.64) implies that $\pi^*_\alpha$ is a connected policy with a

threshold to be labelled as $T_\alpha(i)$ where data compressor 1 is used if $(w,i)$ satisfy $w \leq T_\alpha(i)$ and data compressor 2 is used otherwise. In addition, $T_\alpha(i)$ is a constant for $i \in (0,I_1]$ and $T_\alpha(0) \geq T_\alpha(I_1)$.

<u>Q.E.D</u>

Thus, the α-optimal policy is a connected policy specified by a threshold function $T_\alpha(i)$. The following lemma proves that $T_\alpha(i)$ is upper bounded for all $\alpha \in (\alpha^*,1)$ where $\alpha^* \in (0,1)$.

<u>Lemma 5.4</u>

$T_\alpha(i)$, which is specified in Lemma 5.3, is bounded for all $\alpha \in (\alpha^*,1)$ for some $\alpha^* \in (0,1)$.

<u>Proof</u>

The proof follows from a generalization of the proof of Lemma 3.6.

<u>Q.E.D</u>

Assumption 3.2 is now verified in the following lemma.

<u>Lemma 5.5</u>

Assumption 3.2 is valid for the Markov decision process $(x_n,a_n)$ under consideration with costs $c'(x,a)$ given by (5.41).

<u>Proof</u>

In this proof, the interarrival time is disregarded in the

description of the state, since by Lemma 5.3 $V_\alpha(x,\pi_\alpha^*)$ is independent of the interarrival time. Now consider the following two cases; $T_\alpha(I_1) \geq 0$ and $T_\alpha(I_1) = 0^-$. For the first case of $T_\alpha(I_1) \geq 0$, let $R_1$ and $R_2$ be the disjoint regions defined by

$$R_1 = \{(w,i) \in ([0,\infty) \times [0]) \cup ([0] \times [0,I_1])\} \tag{5.65}$$

and

$$R_2 = \{(w,i) \in ((0,\infty) \times (0,I_1])\} . \tag{5.66}$$

Clearly, if $(w_n,i_n) \in R_1$, then $(w_{n+1},i_{n+1}) \in R_1$. This implies that if $(w_0,i_0) \in R_1$, then for all $n \geq 0$, $(w_n,i_n) \in R_1$ and

$$c'((w_n,i_n),a_n) = c'((w_n,0),a_n) \quad . \tag{5.67}$$

Furthermore, if $(w_0,i_0) \in R_1$ the sequence $\{a_n\}$ is unaltered if the policy $\pi_\alpha^*$ with threshold function $T_\alpha(i)$ is replaced with the policy $\pi_\alpha$ with threshold $T_\alpha(0)$ for all $i \in [0,I_1]$. Hence, the $\alpha$-optimal discounted cost $V_\alpha((w,i),\pi_\alpha^*)$ for $(w,i) \in R_1$ is given by

$$V((w,i),\pi_\alpha^*) = \limsup_{N\to\infty} E_{\pi_\alpha}\{\sum_{n=0}^{N} \alpha^n c'((w_n,0),a_n) | w_0 = w,\ i_0 = i\} \quad . \tag{5.68}$$

Substituting the value of $c'((w,0),a)$ from (5.39) into (5.68), $V_\alpha((w,i),\pi_\alpha^*)$ satisfies

$$V_\alpha((w,i),\pi_\alpha^*) - \frac{(M-1)}{1-\alpha} t_1 = \limsup_{N\to\infty} E_{\pi_\alpha}\{\sum_{n=0}^{N} \alpha^n (w_n + t_{a_n} + \mu D_{a_n}) | w_0 = w,\ i_0 = i\}. \tag{5.69}$$

Comparing the right side of (5.69) to the $\alpha$-optimal discounted cost for a single channel in Section 3.4, it is observed that the right side of (5.69) is equal to the $\alpha$-optimal discounted cost for the single channel. Now from Lemma 3.9 there exists an $\alpha_1^* \in (0,1)$ such that for all $\alpha \in (\alpha_1^*,1)$ and for all $(w,i) \in R_1$, a function $L_1(w)$ exists that satisfies

$$|V_\alpha((w,i),\pi_\alpha^*) - V_\alpha((0,I_1),\pi_\alpha^*)| < L_1(w) \quad . \tag{5.70}$$

Now from the proof of Lemma 5.3, $V_\alpha((w,i),\pi_\alpha^*)$ is shown to be a nonincreasing function of $i \in [0,I_1]$ and a nondecreasing function of $w \in [0,\infty]$. Thus, (5.70) implies that for all $(w,i) \in ([0,\infty) \times [0,I_1])$

$$|V_\alpha((w,i),\pi_\alpha^*) - V_\alpha((0,I_1),\pi_\alpha^*)| \leq L_1(w) \quad . \tag{5.71}$$

Now consider the second case of $T_\alpha(I_1) = 0^-$. For this case let $R_3$ and $R_4$ be the regions defined by

$$R_3 = ([0,\infty) \times (0,I_1]) \tag{5.72}$$

and

$$R_4 = ([0,\infty) \times [0]) \quad . \tag{5.73}$$

So, if $(w_n,i_n) \in R_3$, then $(w_{n+1},i_{n+1}) \in R_3$. This implies that if $(w_0,i_0) \in R_3$, then, for all $n \geq 0$, $(w_n,i_n) \in R_3$, $a_n = 2$, and

$$c'((w_n,i_n),a_n) = c'((w_n,I_1),2) \quad . \tag{5.74}$$

So, for $(w_n,i_n) \in R_3$ the $\alpha$-optimal discounted cost $V_\alpha((w,i),\pi_\alpha^*)$ is given by

$$V_\alpha((w,i),\pi_\alpha^*) = \limsup_{N\to\infty} E_{\pi_\alpha^*}\{ \sum_{n=0}^{N} \alpha^n \, c'((w_n, I_1),2) | w_0 = w, i_0 = I_1\}$$

$$= \limsup_{N\to\infty} E_{\pi_\alpha^*}\{ \sum_{n=0}^{N} \alpha^n (w_n + t_2 + \mu D_a) | w_0 = w, i_0 = I_1)$$

$$+ (M-1)\, t_2/(1-\alpha) \qquad (5.75)$$

where the sequence of waiting times $\{w_n\}$ represents the sequence of waiting times of M/D/1 queueing system with serve duration of $t_2$ and mean interarrival time of $\lambda^{-1}$. Now to find a bound for $V_\alpha((w,i),\pi_\alpha^*) - V_\alpha((0,I_1),\pi_\alpha^*)$, $(w,i) \in ([0,\infty) \times [0,I_1])$, consider the policy $\pi_\alpha$ which always uses data compressor 2. For $(w,i) \in R_3$ the $\alpha$-discounted cost is the same under policy $\pi_\alpha$ or $\pi_\alpha^*$ as shown in (5.75). For $(w,i) \in R_4$ the $\alpha$-discounted cost is higher under policy $\pi_\alpha$ than under policy $\pi_\alpha^*$, since $\pi_\alpha^*$ *is the* $\alpha$*-optimal policy. Thus,* $V_\alpha((w,i),\pi_\alpha^*) - V_\alpha((0,I_1),\pi_\alpha^*)$ is upper bounded by

$$\Delta V_\alpha(w,i) = V_\alpha((w,i),\pi_\alpha) - V_\alpha((0,I_1),\pi_\alpha)$$

$$\geq V_\alpha((w,i),\pi_\alpha^*) - V_\alpha((0,I_1),\pi_\alpha^*) \geq 0 \qquad . \qquad (5.76)$$

To compute $\Delta V_\alpha(w,i)$, consider the sequences of waiting times out of a M/D/1 queueing system given by

$$w_{n+1}^{(i)} = \max(0, w_n^{(i)} + t_2 - \tau_{n+1}) \qquad i = 1, 2 \quad (5.77)$$

where $w_0^1 = 0$, $w_0^2 = w$ and $\{\tau_n\}$ is a sequence of independent and identically distributed exponential random variables with mean $\lambda^{-1}$. Thus, from (5.77) $w \geq w_n^{(2)} - w_n^{(1)} \geq 0$. Furthermore, if $w_m^{(2)} = 0$, then

$w_n^{(2)} = w_n^{(1)}$ for $n \geq m$. So, using these relations for $w_n^{(1)}$ and the expression for $c'((w,i),0)$, $\Delta V_\alpha(w,i)$ is bounded by

$$\Delta V_\alpha(w,i) \leq E_{\pi_\alpha} \sum_{n=0}^{N_w-1} \alpha^n \{w_n^{(2)} - w_n^{(1)} + (M-1)\ (t_1 - t_2)\}$$

$$\leq \{w + (M-1)\ (t_1 - t_2)\} \ E_{\pi_\alpha}(N_w) \tag{5.78}$$

where

$$N_w = \inf\{n: \ w_n^{(2)} = 0\} \quad . \tag{5.79}$$

Setting $z = 0$ in (3.152), $E_{\pi_\alpha}(N_w)$ is given by

$$E_{\pi_\alpha}(N_w) = w + \lambda^{-1} \quad . \tag{5.80}$$

Hence using (5.76), (5.78) and (5.80), for $(w,i) \in ([0,\infty) \times [0,I_1])$

$$V_\alpha((w,i),\pi_\alpha^*) - V_\alpha((0,I_1),\pi_\alpha^*) \leq [w + (M-1)\ (t_1 - t_2)]\ (w + \lambda^{-1}) \ . \tag{5.81}$$

Therefore, by (5.71) and (5.81), for all $\alpha \in (\alpha_1^*,1)$, $|V_\alpha((w,i),\pi_\alpha^*) - V_\alpha([0,I_1],\pi_\alpha^*)|$ is bounded by the maximum of $L_1(w)$ and the right side of (5.81).

Q.E.D

Assumption 3.3 is now verified in the following lemma.

**Lemma 5.6**

For the Markov decision process $(x_n,a_n)$ under consideration with cost function $c'(x,a)$ given by (5.39), there exists an increasing

sequence $\{a_n\}$, $\alpha_n \in (0,1)$, such that $\alpha_n \uparrow 1$, either $T^*_{\alpha_n}(I_1) = 0^-$ or $T^*_{\alpha_n}(I_1) \geq 0$ for all $n > 0$ and $\lim_{n\to\infty} T_n$ exists where

$$T_n = \begin{cases} 0^- & , \text{ if } T^*_{\alpha_m}(I_1) = 0^- \quad \text{for all } m > 0 \\ T^*_{\alpha_n}(0) & , \text{ if } T^*_{\alpha_m}(I_1) \geq 0 \quad \text{for all } m > 0 \end{cases} \quad . \tag{5.82}$$

Furthermore, the connected policy $\pi'$ with threshold $T'(i)$ given by

$$T'(i) = \lim_{n\to\infty} T_n \;, \quad i \in [0,I_1] \tag{5.83}$$

satisfies

$$\Phi'(x,\pi') = \lim_{n\to\infty} \sup \, (1 - \alpha_n)\, V_{\alpha_n}(x,\pi^*_{\alpha_n}) \tag{5.84}$$

for all $x \in X$.

Proof

Since Lemma 5.4 showed the existence of an $\alpha^* \in (0.1)$ such that, for all $\alpha \in (\alpha^*,1)$, $T^*_\alpha(i)$ is bounded, then there exists an increasing sequence $\{\alpha_n\}$, $\alpha_n \in (0,1)$, such that $\alpha_n \uparrow 1$ and either $T^*_{\alpha_n}(I_1) = 0^-$ for all n or $T^*_{\alpha_n}(I_1) \geq 0$ for all n with the $\lim_{n\to\infty} T^*_{\alpha_n}(0)$ well defined. Thus, it is clear that $T'(\cdot)$ defined by (5.84) exists, and the connected policy $\pi'$ with threshold $T'(i)$ is well specified.

Consider the case of $T^*_{\alpha_n}(I_1) \geq 0$ for all $n > 0$. Then $T'(i)$ is given by

$$T'(i) = \lim_{n\to\infty} T^*_{\alpha_n}(0) \quad . \tag{5.85}$$

Using (5.69) for $x \in (R_1 \times [0,\infty))$,

$$(1-\alpha_n)V_{\alpha_n}(x,\pi^*_\alpha) - (M-1)t_1$$

$$= \limsup_{N\to\infty} E_{\pi_{\alpha_n}} \sum_{m=0}^{N} \alpha^m(w_m + t_{a_m} + \mu D_{a_m})|x_0 = x\}(1-\alpha_n) \tag{5.86}$$

where $\pi_\alpha$ is the connected policy with threshold function $T_\alpha(i) = T^*_\alpha(0)$ for $i \in [0,I_1]$ and $R_1$ is given by (5.65). In a similar manner to (5.71), it is clear that for $x \in (R_1 \times [0,\infty))$

$$\Phi'(x,\pi') - (M-1)t_1 = \limsup_{N\to\infty} N^{-1} E_{\pi'}\{\sum_{n=0}^{N-1} (w_n + t_{a_n} + \mu D_{a_n})|x_0 = x\} \quad . \tag{5.87}$$

Since the right side of (5.86) represents the $\alpha$-optimal cost for the single channel case which by Lemma 3.11 converges to the right side of (5.87) as $n \to \infty$, then for $x \in (R_1 \times [0,\infty))$

$$\Phi'(x,\pi') = \limsup_{n\to\infty} (1 - \alpha_n) V_{\alpha_n}(x,\pi^*_{\alpha_n}) \quad . \tag{5.88}$$

Now it is easily shown that $\Phi'(x,\pi')$ is a constant for all $x \in X$. Hence using Lemma 5.5 and (5.88) for all $x \in X$

$$\begin{aligned}
&|\Phi'(x,\pi') - \limsup_{n\to\infty} (1 - \alpha_n) V_{\alpha_n}(x,\pi^*_{\alpha_n})| \\
&\quad \le |\Phi'((0,I_1,0),\pi') - \limsup_{n\to\infty} (1 - \alpha_n) V_{\alpha_n}([0,I_1,0),\pi^*_{\alpha_n})| \\
&\quad + \limsup_{n\to\infty} (1 - \alpha_n)|V_{\alpha_n}(x,\pi^*_{\alpha_n}) - V_{\alpha_n}([0,I_1,0),\pi^*_{\alpha_n})| \\
&\quad = 0
\end{aligned} \tag{5.89}$$

Now consider the case of $T^*_{\alpha_n}(I_1) = 0^-$ for all $n \ge 0$. Then $T'(i) = 0^-$ for $i \in [0,I_1]$. Clearly from (5.75) for $x \in (R_3 \times [0,\infty))$

where $R_3$ is given by (5.72) and for $\pi'$ a connected policy with threshold $T'(i)$,

$$\Phi'(x,\pi') = \limsup_{n\to\infty} (1 - \alpha_n) V_{\alpha_n}(x,\pi')$$

$$= \limsup_{n\to\infty} (1 - \alpha_n) V_{\alpha_n}(x,\pi^*_{\alpha_n}) \quad . \tag{5.90}$$

It is clear that $\Phi'(x,\pi')$ is a constant for all $x \in X$. So, using Lemma 5.5 and (5.90), in a manner similar to (5.89), it is found that for $x \in X$

$$\Phi'(x,\pi') = \limsup_{n\to\infty} (1 - \alpha_n) V_{\alpha_n}(x,\pi^*_{\alpha_n}) \quad . \tag{5.91}$$

<u>Q.E.D</u>

Finally, using Theorem 3.2, the average optimal policy is established.

<u>Lemma 5.7</u>

The policy $\pi'$ which satisfies for all $x \in X$

$$\Phi'(x) = \Phi'(x,\pi') = \Phi'((0,I_1,0),\pi') \tag{5.92}$$

is given by the policy described in Lemma 5.6.

**Proof**

Since by Lemmas 5.2, 5.5, and 5.6, Assumptions 3.1, 3.2, and 3.3 are valid, then the results stated in the theorem follow by Theorem 3.2.

Q.E.D

Now to verify that $\pi'$ is also a policy $\pi^*$ which satisfies (5.38), it is required to show, as was discussed earlier, that for all $x \in X$ $\Phi'(x)$ is equal to $\Phi(x,\pi')$. The following theorem yields the required result.

**Theorem 5.1**

Let the average cost $\Phi(x)$ be given by (5.34) and let $\Phi'(x)$ be given by (5.92) where the policy $\pi'$ given in Lemma 5.6 achieves $\Phi'(x)$. Then

$$\Phi(x) = \Phi'(x) = \Phi'((0,I_1,0)) \quad . \tag{5.93}$$

Furthermore, the policy $\pi'$ satisfies (5.38) and is the average optimal policy for the Markov decision problem with costs $c(x,a)$ given by (5.32).

**Proof**

First, it is required to prove that, for all $x \in X$, $\Phi(x,\pi')$ equals $\Phi'(x,\pi')$. Generalizing Lemma 3.10, it is clear that under $\pi'$ the distribution of $x_n = (w_n,i_n,\tau_n)$ has a limiting probability distribution, neglecting $\tau_n$, given by

$$F(W,I) = \lim_{n\to\infty} \Pr(w_n \le W,\ i_n \le I \mid x_0 = x) \tag{5.94}$$

which is independent of $x \in X$. Furthermore, $\lim_{n\to\infty} E(w_n \mid x_0 = x)$ exists, is finite and is also independent of $x \in X$. So, $\phi(x,\pi')$ and $\phi'(x,\pi')$ are given in terms of $F(W,I)$ by

$$\phi(x,\pi') = \int_{\substack{w\in[0,\infty) \\ i\in[0,I_1]}} c((w,i),a(w))\, dF(w,I) \tag{5.95}$$

and

$$\phi'(x,\pi') = \int_{\substack{w\in[0,\infty) \\ i\in[0,I_1]}} c'((w,i),a(w))\, dF(w,i) \tag{5.96}$$

where $a(w)$ is the decision function given by

$$a(w) = \begin{cases} 1, & \text{if} \quad w \le T'(i) \\ 2, & \text{if} \quad w > T'(i) \end{cases}, \tag{5.97}$$

and $T'(i)$ is given by (5.83) and is a constant independent of $i$.

Now consider the case where the threshold $T'(i)$ associated with $\pi'$ is greater than or equal to zero. Then it is clear that no state in the region $R_1 \times [0,\infty)$ communicates with a state in the region $R_2 \times [0,\infty)$ where $R_1$ and $R_2$ are given by (5.65) and (5.66), respectively. Since from (5.94) the limiting probability distribution is independent of the initial state, then all states in $R_2 \times [0,\infty)$ are transient states and

$$\lim_{n\to\infty} \Pr\{x_n \in (R_2 \times [0,\infty))\} = 0 \quad . \tag{5.98}$$

But for all $(w,i) \in R_1$

$$c((w,i),a(w)) = c'((w,i),a(w)) \quad . \tag{5.99}$$

Therefore, using (5.98) and (5.99) in (5.95) and (5.97) results in

$$\Phi(x,\pi') = \Phi'(x,\pi') = \Phi((0,I_1,0)) \tag{5.100}$$

for all $x \in X$.

For the case where the threshold $T'(i)$ associated with $\pi'$ is $0^-$, it is clear that no state in the region $([0,\infty) \times [I_1] \times [0,\infty))$ communicates with a state in the region $([0,\infty) \times [0,I_1) \times [0,\infty))$. Thus, in a similar manner to the case of $T'(i) \geq 0$,

$$\lim_{n\to\infty} \Pr\{x_n \in ([0,\infty) \times [0,I_1) \times [0,\infty))\} = 0 \tag{5.101}$$

and for all $(w,i) \in ([0,\infty) \times [I_1])$

$$c((w,i),a(w)) = c'((w,i),a(w)) \quad . \tag{5.102}$$

Hence substituting (5.101) and (5.102) into (5.95) and (5.96), the following result is obtained:

$$\Phi(x,\pi') = \Phi(x,\pi') = \Phi((0,I_1,0)) \tag{5.103}$$

for all $x \in X$.

Now from (5.42) for all $x \in X$, $\Phi'(x) \leq \Phi(x)$, and from (5.100) and (5.103) for all $x \in X$,

$$\Phi'(x) = \Phi'(x,\pi') = \Phi(x,\pi') \quad . \tag{5.104}$$

Therefore, for all $x \in X$

$$\Phi(x) = \Phi(x,\pi') = \Phi(0,I_1,0) \quad . \tag{5.105}$$

Q.E.D

Thus, from (5.37) it is clear that $\pi'$ given in Lemma 5.6 satisfies (5.3) and so,

$$L^2(\{D_k\}) = L^2_{\pi'}(\{D_k\}) \quad . \tag{5.106}$$

Hence, equation (3.16) is a valid expression for the delay distortion relation $\gamma_2^M(D)$. Since $\pi'$ is a connected policy with threshold $T'(i) = T'$ applied to the waiting time, then it is clear that

$$L^2(\{D_k\}) = \min_{\pi\epsilon\Pi} L^2_\pi(\{D_k\})$$

$$= \min_{T\epsilon[[0,\infty)\cup(0^-)]} L^2_{\pi_T}(\{D_k\}) \tag{5.107}$$

where $\pi_T$ is the connected policy with threshold T. In evaluating $L^2_{\pi_T}(\{D_k\})$, the case of a threshold of $0^-$ corresponds to the case of $D_1 = D_2$ and a policy with an arbitrary positive threshold. Thus, substituting (5.107) into (3.16) results in

$$\gamma_2^M(D^*) = \min_{\substack{\{D_k \geq 0, k=1,2\} \\ \{T \geq 0\}}} \{L^2_{\pi_T}(\{D_k\})\} - \mu D^* \tag{5.108}$$

where $\{D_k^*\}$ and $T^*$ minimize $L^2_{\pi_T}(\{D_k\})$ and $D^*$ is the average distortion associated with $\{D_k^*\}$ and $T^*$. Hence, $\gamma_2^M(D)$ for the tandem channel case is computed by finding the minimum of a functional in much the same manner as for the single channel case described in Section 4.1. The

functional $L^2_{\pi_T}(\{D_k\})$, $T \geq 0$, is readily evaluated in terms of expressions derived in Section 4.3, and the following lemma describes this relationship.

Lemma 5.8

For $D_1 \geq D_2$ and $T \geq 0$,

$$L^2_{\pi_T}(\{D_k\}) = \int_0^\infty W\, dF(W) + (t_1 + \mu D_1)\, F(T) + (t_2 + \mu D_2)\,(1 - F(T)) + (M-1)t_1 \tag{5.109}$$

where $t_k$ is given by (5.5) and $F(W)$ is the probability distribution given in Theorem 4.2.

Proof

The expressions (5.95), (5.96) and (5.99) used in the proof of Theorem 5.1 are all valid for any connected policy with a non-negative threshold. Also, for the case of a positive threshold, it is clear that $c((0,i),a(0))$ in the integral in (5.95) is a constant independent of $i$. So,

$$L^2_{\pi_T}(\{D_k\}) = \Phi((0,I_1,0),\pi_T) = \int_{w\varepsilon[0,\infty)} c((w,0),a(w))\, dF(w) \tag{5.110}$$

where $F(W)$, which is given in Theorem 4.2, is the limiting distribution of the waiting time in the buffer associated with the first channel.

Now substituting the value of $c((w,0),a(w))$ into (5.110) the expression (5.109) results.

Q.E.D

Thus, the delay distortion relationship $\gamma_2^M(D)$ for a tandem channel network is calculated in the same manner which was used for calculating the delay distortion relationship for a single channel network.

## 5.2 Properties of $\gamma_2^M(D)$

The delay distrotion relationship $\gamma_2^M(D)$ for the tandem channel network defined by (3.13) has the properties which are described in Section 4.2 for $\gamma_2^1(D)$. That is, $\gamma_2^M(D)$ is a nonincreasing function of distortion D, and $D^2_{min}$ defined in (4.10) is given by

$$D^2_{min} = \hat{r}^{-1}(\lambda^{-1} C - p_2) \quad , \tag{5.111}$$

and is independent of the number of channels.

An additional property which is of interest is associated with the behavior of $\gamma_2^M(D)$ as a function of the number channels M for a fixed D. The following theorem indicates that the advantage of using an adaptive data compression decays as the number of channels in tandem increases.

**Theorem 5.2**

For the source and tandem channel network described in Section 3.1, let $C_m = C$ for all m and let $D > D^2_{min}$. Then

$$p_2\, C^{-1} + (\gamma_1^M(D) - \gamma_2^M(D)) \leq (\gamma_1^{M-1}(D) - \gamma_2^{M-1}(D)) \qquad (5.112)$$

where $p_2$ is the associated protocol information.

**Proof**

Suppose that policy $\pi$ and distortion values $D_1$ and $D_2$, $D_1 \leq D_2$, achieve $\gamma_2^M(D)$. From (5.108) it is clear that $\pi$ must be a connected policy and $D_1 \leq D$. Let $T_\pi^2(\{D_k\})_{M-1}$ be the delay through a tandem network of M-1 channels when policy $\pi$ is used with data compressors with associated distortion levels of $D_1$ and $D_2$. From the expression for $T_\pi^2(\{D_k\})_M$ given by (3.11), it is clear that

$$\gamma_2^M(D) - T_\pi^2(\{D_k\})_{M-1} = (\hat{r}(D_1) + p_2)C^{-1}. \qquad (5.113)$$

Since the policy $\pi$ together with distortion values $D_1$ and $D_2$ achieve an average distortion less than or equal to D, then

$$T_\pi^2(\{D_k\})_{M-1} \geq \gamma_2^{M-1}(D) \quad . \qquad (5.114)$$

Thus,

$$\gamma_2^M(D) - \gamma_2^{M-1}(D) \geq [\hat{r}(D) + p_2]C^{-1} \qquad (5.115)$$

where $\hat{r}(D)$ is assumed to be a nonincreasing function of D.

Now from (3.10),

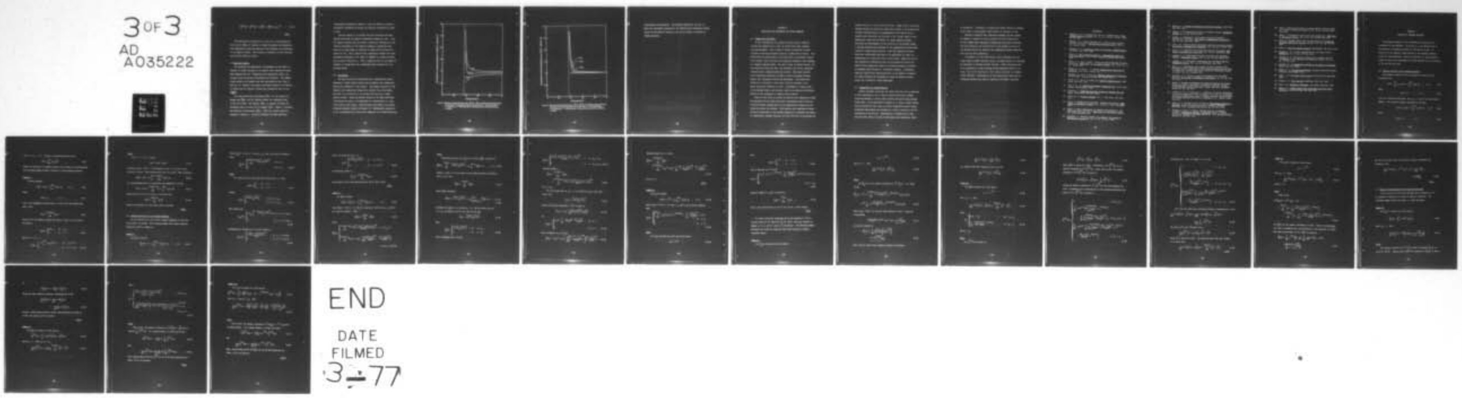
$$\gamma_1^M(D) - \gamma_1^{M-1}(D) = \hat{r}(D)\, C^{-1} \quad . \qquad (5.116)$$

So, using (5.115) and (5.116), one obtains

AD-A035 222 CALIFORNIA UNIV LOS ANGELES SCHOOL OF ENGINEERING A--ETC F/G 17/2
SOURCE CODING AND ADAPTIVE DATA COMPRESSION FOR COMMUNICATION N--ETC(U)
DEC 76 H S NUSSBAUM N00014-75-C-0609

UNCLASSIFIED UCLA-ENG-76121 NL

3 OF 3
AD
A035222



END
DATE
FILMED
3-77

$$\{\gamma_1^{M-1}(D) - \gamma_2^{M-1}(D)\} - \{\gamma_1^M(D) - \gamma_2^M(D)\} \geq p_2 \, C^{-1} \quad . \tag{5.117}$$

Q.E.D

The implication of Theorem 5.2 is that for a fixed distortion level as the number of channels in tandem increases, the nonadaptive data compression scheme approaches and then surpasses the performance of the adaptive scheme. This property is exhibited in the following section with numerical results.

## 5.3 Numerical Results

To illustrate the degradation in performance as the number of channels in tandem increases of an adaptive data compression scheme when compared with the nonadaptive data compression scheme, the relevent delay distortion relationships are evaluated. The communication system to be examined uses the same models as were used in Section 4.5. However, the channel considered consists of M channels in tandem with all channels transmitting information bits at rate $C[\frac{\text{nats}}{\text{sec.}}]$.

The delay distortion relationship $\gamma_1^M(D)$, for the nonadaptive scheme, and $\gamma_2^M(D)$, for the adaptive scheme, are evaluated using (3.10) and (5.110). The quantity $\gamma_2^M(D)$ is computed following the procedures used in Section 4.5 to compute $\gamma_2^1(D)$. Figure 5.1 presents the difference relationships $\gamma_1^M(D) - \gamma_2^M(D)$ for various numbers of channels in tandem, M. Figure 5.2 presents the same difference

relationships presented in Figure 5.1 with the effects of protocol information eliminated by setting the protocol information $p_2$ equal to zero.

From the figures it is evident that as M increases the gains derived from using the adaptive compression schemes are lost. Also, the figures indicate that the effects of protocol information on the relative performance of the adaptive scheme are significant when there are a large number of channels in tandem and the distortion levels are low. Hence, adaptive data compression schemes lose their performance advantage as the number of channels in tandem increases as was shown in Theorem 5.2. This is expected since as the number of channels in tandem grows the transmission delay prevails over the queueing delays.

## 5.4 Conclusions

The delay distortion relationship for a communication system employing a tandem channel network and an adaptive data compression scheme was considered in this chapter. The optimal structure of the adaptive data compression scheme which achieves this relationship was shown to be identical to the optimal structure for a single channel network. Using the optimal structure, the delay distortion relationship was shown to be specified by a minimization of a functional over a vector space. This functional was shown to be readily evaluated using the results of Chapter IV. Subsequently, properties of the relationship were reviewed and examples of the delay distortion

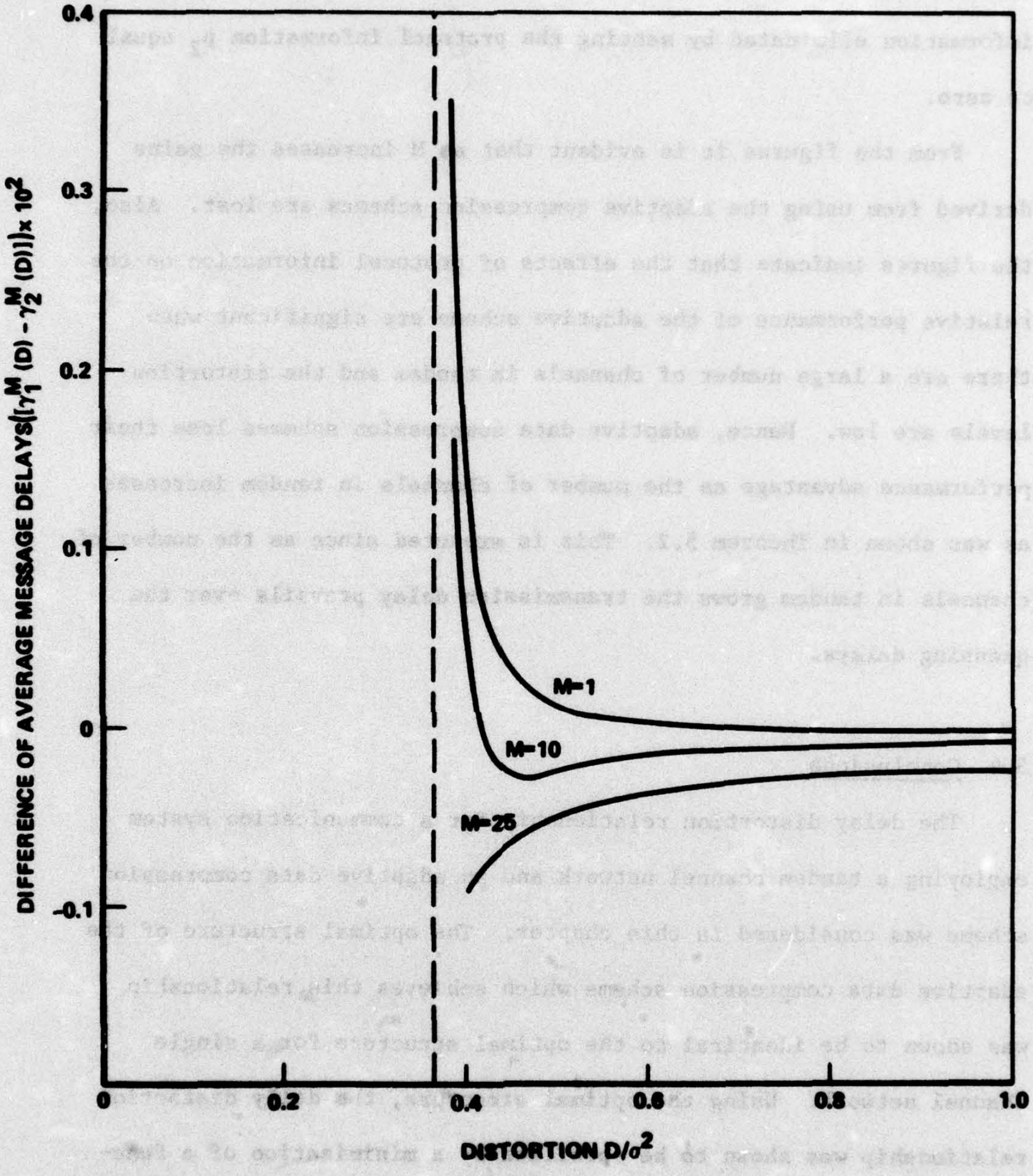



**Figure 5.1. Difference of Delay Distortion Curves, $\gamma_1^M(D) - \gamma_2^M(D)$, Including the Effects of Protocol Information for Tandem Channel Systems, Gaussian Source, $\nu$ = 100 [Letters/Mess.], C = [Knats/Sec.], $\lambda$ = 1000 [Mess./Sec.], with Parameter M the Number of Channels in Tandem**

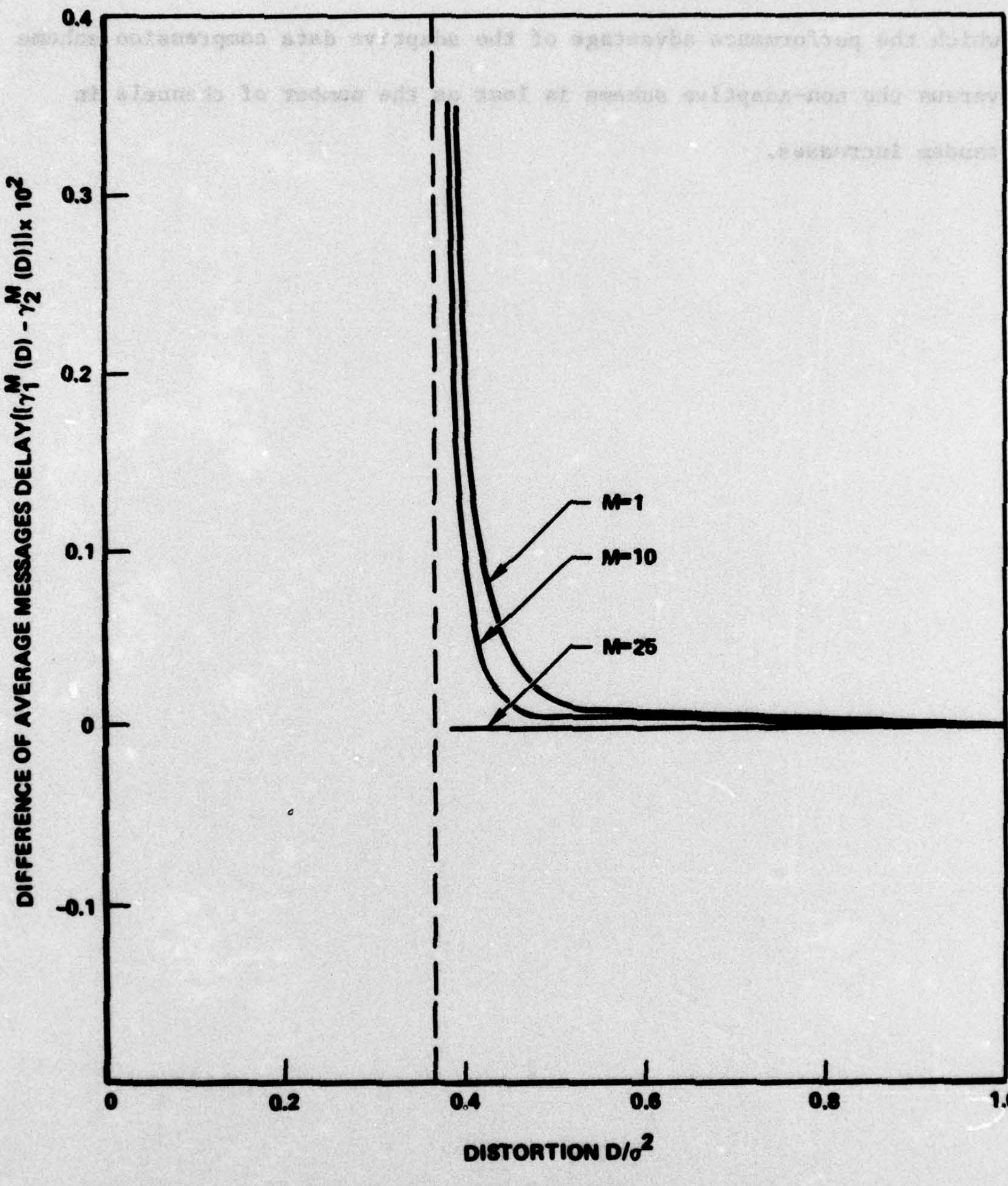



**Figure 5.2. Difference of Delay Distortion Curves, $\gamma_1^M$ (D) - $\gamma_2^M$ (D), Excluding the Effects of Protocol Information for Tandem Channel Systems, Gaussian Source, $\nu$ = 100 [Letters/Mess.], C = 50 [Knats/Sec.], $\lambda$ = 1000 [Mess./Sec.], with Parameter M the Number of Channels in Tandem**

relationship were presented. The examples demonstrate the rate at which the performance advantage of the adaptive data compression scheme versus the non-adaptive scheme is lost as the number of channels in tandem increases.

## CHAPTER VI

## CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

### 6.1 Summary and Conclusion

In this dissertation the coding of sources under a fidelity criteria was examined for a class of sources which emit randomly occurring messages. Such a class of sources is employed to model information carrying processes entering a communication network. They differ from the normal models of sources found in information theoretical studies, since the latter are generally assumed to emit messages on a regular temproal basis. For such a class of network sources, the rate distortion function, R(D), was evaluated, and source coding and converse source coding theorems were proved. From these theorems a new operational definition of R(D) in terms of message queueing delays, and transmission delays was determined. This operational definition is observed to constitute a natural extension of the usual operational definition of R(D). Furthermore, it relates R(D) to the message delay in the network, which is an important performance measure in the evaluation of communication networks.

Then for this class of sources an adaptive data compression scheme was presented and the delay distortion relationship, which forms the trade-off between message delay in the communication network and the distortion level, was studied. This adaptive data compression scheme utilizes observations of the network congestion to determine the amount of compression a message receives, with the objective of minimizing the

message delay for a given distortion level. Using results from Markov decision theory and extensions, which were derived in the dissertation, of Markov decision theory to nondenumerable state spaces and unbounded costs functions, the structure of the optimal adaptive data compression scheme was determined for tandem channel networks. Following the establishment of the structure of the optimal scheme, a queueing analysis was performed for the communication system which utilizes a tandem channel network. From this analysis the delay distortion relationship was shown to be expressible in terms of a minimization of a functional over a vector space. Numerical results were presented to demonstrate that for low distortion levels the adaptive data compression scheme is an effective means of reducing the message delay obtained by nonadaptive schemes. Furthermore, numerical results have demonstrated that as the number of channels in tandem increased the advantages of using the adaptive scheme versus the nonadaptive scheme diminished.

## 6.2 Suggestions for Future Research

Several possible directions for future work will now be indicated. In this dissertation, for a class of sources which emit randomly occurring messages, a source coding theorem was proved using message block codes. As was mentioned in Chapter II, a source coding theorem could be proved for this class of sources using block source coding procedures which employ the encoding of a block of duration T of the realization of the source. Investigation of message block codes and time block codes, in terms of performance and complexity, would

be worthwhile. Furthermore, extending the coding theorem for message block codes to convolutional codes would be of interest as well.

Related to adaptive data compression schemes, further studies are needed to extend the results presented in this dissertation to more complex data compression schemes and more complex networks. One such problem is the determination of the optimal structure of the decision policy for adaptive data compression schemes employing more than two data compressors.

Furthermore, in this dissertation, novel techniques were derived, based on Markov decision theory, to answer questions concerning the existence of optimal decision policies. These techniques seem to be applicable to other queueing systems employing adaptive controls such as the multitude of flow control procedures in communication networks. Investigation of such applications employing the techniques derived here would be important and highly rewarding.

## REFERENCES


1. Shannon, C. E. "A Mathematical Theory of Communication," Bell Systems Technical Journal, Vol. 27, pp. 379-423, 623-656, July, October 1948.

2. Shannon, C. E. "Coding Theorems for a Discrete Source with a Fidelity Criterion," IRE National Convention Record, Part 4, pp. 142-163, 1959.

3. Gallager, R. G. Information Theory and Reliable Communications, John Wiley, New York, 1968.

4. Berger, T. Rate Distortion Theory: A Mathematical Basis for Data Compression, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

5. Kirch, S. I. and T. Berger. "Coding for Delay-Dependent Fidelity Criterion," IEEE Trans. on Information Theory, Vol. IT-20, pp. 77-85, January 1974.

6. Davies, D. W. and D. L. A. Barber. Communication Networks for Computers, John Wiley, London, 1973.

7. Abramson, N. and F. F. Kuo, ed. Computer-Communication Networks, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

8. Green, P. E. and R. W. Lucky, ed. Computer Communications, IEEE Press, New York, 1975.

9. Chu, W. W., ed. Advances in Computer Communications, Artech House, Dedham, Massachusetts, 1974.

10. Kleinrock, L. Communication Nets; Stochastic Message Flow and Delay, McGraw-Hill, New York, 1964.

11. Kleinrock, L. Queueing Systems, Vol. 2, John Wiley, New York, 1976.

12. Rubin, I. "Communication Networks: Message Path Delays," IEEE Trans. on Information Theory, Vol. IT-20, pp. 738-45, November 1974.

13. Rubin, I. "Data Compression for Communication Networks: The Delay-Distortion Function," IEEE Trans. on Information Theory, Vol. IT-22, November 1976.

14. Blackwell, D. "Discrete Dynamic Programming," The Annals of Mathematical Statistics, Vol. 33, pp. 719-726, 1962.

15. Howard, R. A. Dynanic Programming and Markov Processes, John Wiley, New York, 1960.

16. Derman, C. "On Sequential Decisions and Markov Chains," Management Science, pp. 16-24, October 1962.

17. Derman, C. "Denumerable State Markov Decision Processes - Average Cost Criterion," The Annals of Mathematical Statistics, Vol. 37, pp. 1545-1554, 1966.

18. Ross, S. M. "Non-Discounted Denumerable Markovian Decision Models," The Annals of Mathematicsl Statistics, Vol. 39, pp. 412-423, 1968.

19. Ross, S. A. "Arbitrary State Markovian Decision Processes," The Annals of Mathematical Statistics, Vol. 39, pp. 2118-2122, 1968.

20. Lippman, S. A. "Semi-Markov Decision Processes with Unbounded Rewards," Management Science, Vol. 19, pp. 717-731, March 1973.

21. Lippman, S. A. "On Dynamic Programming with Unbounded Rewards," Management Science, Vol. 21, pp. 1225-1233, July 1975.

22. Reed, F. C. "Denumerable State Markov Decision Processes with Unbounded Costs," Department of Operations Research and Department of Statistics, Stanford University, Stanford, California, Technical Report No. 157, November 1973.

23. Harrison, J. M. "Discrete Dynamic Programming with Unbounded Rewards," The Annals of Mathematical Statistics, Vol. 43, pp. 636-644, 1972.

24. Awate, P. "Dynamic Programming with Negative Rewards and Average Reward Criterion," Department of Operations Research, College of Engineering, Cornell University, Ithaca, New York, Technical Report No. 251, May 1975.

25. Crabill, T. B., D. Gross and M. J. Magazine. "A Survey of Research on Optimal Design and Control of Queues," Institute for Management Science and Engineering, School of Engineering and Applied Science, The George Washington University, Washington, D.C., Serial T-280, June 1973.

26. Sobel, M. J. "Optimal Control of Queues," Mathematical Methods in Queueing, Western Michigan University, Kalamazoo, Michigan, May 10-12, 1973, Spring-Verlag, Berlin, pp. 231-261, 1974.

27. Stidham, S. and N. U. Prabhu. "Optimal Control of Queueing Systems," Mathematical Methods in Queueing, Western Michigan University, Kalamazoo, Michigan, May 10-12, 1973, Springer-Verlag, Berlin, pp. 264-294, 1974.

28. Shaw, L. "Busy Period Control of Queues Based on Waiting Times at Arrivals," Operations Research, Vol. 24, pp. 543-563, May-June 1976.

29. Rubin, I. "Information Rates for Poisson Sequences," IEEE Trans. on Information Theory, Vol. IT-19, pp. 283-294, May 1973.

30. Smith, W. "Renewal Theory and Its Ramifications," Journal of the Royal Statistical Society, Series B, Vol. 20, pp. 243-302, 1958.

31. Rudin, W. Real and Complex Analysis, McGraw-Hill, New York, 1966.

32. Kingman, J. F. C. "Some Inequalities for the Queue GI/G/1," Biometrika, Vol. 49, pp. 315-324, 1962.

33. Lindley, D. V. "The Theory of Queues with a Single Server," Proceedings of the Cambridge Philosophical Society, Vol. 48, pp. 277-289, 1952.

34. Kinchine, A. Y. Mathematical Methods in the Theory of Queueing, Hafner, New York, 1969.

35. Widder, D. V. The Laplace Transform, Princeton University Press, Princeton, New Jersey, 1941.

36. Cohen, J. W. The Single Server Queue, North-Holland, Amsterdam, 1969.

37. Royden, H. L. Real Analysis, MacMillan, New York, 1968.

38. Doob, J. B. Stochastic Processes, John Wiley, New York, 1953.

39. Chung, K. L. Markov Chains with Stationary Transition Probabilities, Springer-Verlag, New York, 1967.

## APPENDIX

## SOLUTION TO INTEGRAL EQUATION

The integral equation (4.24) introduced in Section 4.3 is considered in this Appendix. In Section A.1, the general form of the solution to integral equations of this type is found. In Section A.2, the specific forms related to the integral equation (4.24) are then developed. Finally in Section A.3, these specific forms are used in the calculation of terms required in the evaluation of the mean waiting time.

### A.1 General Solution to the Integral Equation

The integral equation (4.24) can always be written in the following form:

$$G(x) = \sum_{i=1}^{N} a_i \, z_i(x) - b \int_{-\infty}^{x-c} G(y)\, dy \quad , \quad x \geq T \tag{A.1}$$

where

$$G(x) = 0 \quad , \quad x < T \quad , \tag{A.2}$$

$z_i(x)$ are integrable functions, and $a_i$, b, c, and T are real valued numbers. Now consider integral equations of the form

$$G_i(x) = z_i(x) - b \int_{-\infty}^{x-c} G_i(y)\, dy \; , \quad x \geq T \tag{A.3}$$

where

$$G_i(x) = 0 \quad , \quad x < T \tag{A.4}$$

for i = 2, 2, ..., N. Clearly, by substitution into (A.1)

$$G(x) = \sum_{i=1}^{N} a_i \, G_i(x) \quad . \tag{A.5}$$

Hence, the solution to integral equation (A.3) needs to be investigated. The following lemma provides a solution to this integral equation.

<u>Lemma A.1</u>

Let G(x) satisfy

$$G(x) = z(x) - b \int_{-\infty}^{x-c} G(y)\, dy \qquad x \geq T \tag{A.6}$$

where

$$G(x) = 0 \quad , \quad x < T \quad , \tag{A.7}$$

z(x) is an integrable function and b, c, and T are real valued numbers. Then

$$G(x) = \sum_{r=0}^{[(x-T)/c]} d_r(x) \tag{A.8}$$

where [x] is the largest integer less than or equal to x, and $d_r(x)$ is given by

$$d_0(x) = \begin{cases} z(x) \quad , & \text{if} \quad x \geq T \\ 0 \quad , & \text{if} \quad T < x \end{cases} \tag{A.9}$$

and for r = 1, 2, ...

$$d_r(x) = \begin{cases} -\, b \int_{T}^{x-c} d_{r-1}(y)\, dy \quad , & \text{if} \quad x \geq T + rc \\ 0 \quad , & \text{if} \quad x < T + rc \quad . \end{cases} \tag{A.10}$$

**Proof**

For $T \leq x < T + c$, clearly

$$G(x) = z(x) = d_0(x) \qquad \text{(A.11)}$$

satisfies (A.6). Let $\hat{x} > T$ and assume for all $x < \hat{x}$, (A.8) is the solution of (A.6). Then substituting (A.8) into (A.6), $G(\hat{x})$ is given by

$$G(\hat{x}) = z(\hat{x}) - b \int_T^{\hat{x}-c} \sum_{r=0}^{[(y-T)/c]} d_r(y)\, dy \quad . \qquad \text{(A.12)}$$

So, interchanging order of integration and summation in (A.12),

$$G(\hat{x}) = z(\hat{x}) + \sum_{r=0}^{[(\hat{x}-T-c)/c]} \left(- b \int_T^{\hat{x}-c} d_r(y)\, dy\right) \quad . \qquad \text{(A.13)}$$

Using (A.9) and (A.10) in (A.13), it is clear that

$$G(\hat{x}) = \sum_{r=0}^{[(\hat{x}-T)/c]} d_r(\hat{x}) \quad . \qquad \text{(A.14)}$$

Hence, by induction on $\hat{x}$ the result (A.8) is proved.

Q.E.D

## A.2 Specific Solution to the Integral Equation

In the solution of (4.24), three integral equations of the form (A.6) need to be solved. The following lemmas solve these equations using the results of Lemma A.1.

**Lemma A.2**

Let $G_{t_k}^T(x)$ satisfy

$$G_{t_k}^T(x) = 1 - \mu_k \int_{\infty}^{x-t_i} G_{t_i}^T(y)\, dy \quad , \quad x \geq T \qquad \text{(A.15)}$$

where $G_{t_k}^T(x) = 0$ for $x < T$ and $\mu_k$, $t_k$, and T are positive numbers.

Then

$$G_{t_k}^T(x) = \begin{cases} \displaystyle\sum_{r=0}^{[(x-T)/t_k]} \frac{(-\mu_k(x - rt_k))^r}{r!} & , \text{ if } x \geq T \\ 0 & , \text{ if } x < T \quad . \end{cases} \tag{A.16}$$

**Proof**

Let $d_r(x)$ be as in (A.9) and (A.10) with $z(x) = 1$. Then

$$d_0(x) = \begin{cases} 1 & , \text{ if } \quad x \geq T \\ 0 & , \text{ if } \quad x < T \quad . \end{cases} \tag{A.17}$$

Assume

$$d_r(x) = \begin{cases} \dfrac{(-\mu_k(x - rt_k))^r}{r!} & , \text{ if } x \geq T + rt_k \\ 0 & , \text{ if } x < T + rt_k \quad . \end{cases} \tag{A.18}$$

Then from (A.10)

$$d_{r+1}(x) = \begin{cases} -\mu_{k_T} \displaystyle\int_T^{x-t} \frac{(-\mu_k(y - rt_k))^r}{r!} \, dy, & \text{if } \ x \geq T + (r+1)c \\ 0 & , \text{if } \ x < T + (r+1)c \ . \end{cases} \tag{A.19}$$

Performing the integration in (A.19) results in

$$d_{r+1}(x) = \begin{cases} \dfrac{(-\mu_k(x - (r+1)t_k))^{r+1}}{(r+1)!} & , \ \text{if } \ x \geq T + (r+1)c \\ 0 & , \ \text{if } \ x < T + (r+1)c \ . \end{cases} \tag{A.20}$$

Hence, by induction for $r \geq 0$

$$d_r(x) = \begin{cases} \dfrac{(-\mu_k(x - rt_k))^r}{r!} & , \text{ if } \quad x \geq T + rc \\ 0 & , \text{ if } \quad x < T + rc \end{cases} \tag{A.21}$$

and applying Lemma A.1,

$$G_{t_i}^T(x) = \sum_{r=0}^{[(x-T)/t_k]} d_r(x) \tag{A.22}$$

which gives (A.16) upon substitution of (A.21) into (A.22).

<u>Q.E.D</u>

<u>Lemma A.3</u>

Let $G_G^T(x)$ satisfy

$$G_G^T(x) = G_{t_1}^0(x) - \mu_2 \int_T^{x-t_2} G_G^T(y)\, dy \quad , \quad x \geq T \tag{A.23}$$

where $G_G^T(x) = 0$ for $x < T$, $G_{t_1}^0(x)$ is given by (A.16) and $\mu_2$, $t_2$ and T are positive numbers. Then

$$G_G^T(x) = \sum_{r=0}^{[x/t_1]} \hat{G}_r^T(x) \tag{A.24}$$

where

$$\hat{G}_r^T(x) = \begin{cases} \displaystyle\sum_{v=0}^{[(x-rt_1)/t_2]} \frac{(-\mu_1)^r\ (-\mu_2)^v\ (x-rt_1-vt_2)^{r+v}}{(r+v)!} & , \text{ if } rt_1 \geq T \\ \displaystyle\sum_{v=0}^{[(x-T)/t_2]} (-\mu_1)^r\ (-\mu_2)^v \sum_{u=0}^{r} \left\{ \frac{(x-T-vt_2)^{u+v}}{(u+v)!} \quad \frac{(T - ut_1)^{r-u}}{(r-u)!} \right\} & , \text{ if } rt_1 < T. \end{cases} \tag{A.25}$$

**Proof**

Substituting (A.16) for $G^0_{t_1}(x)$ in (A.23), $G^T_G(x)$ is given by

$$G^T_G(x) = \sum_{r=0}^{[x/t_1]} \frac{(-\mu_1(x - rt_1))^r}{r!} - \mu_2 \int_T^{x-t_2} G^T_G(y)\, dy \quad , \quad x \geq T. \quad \text{(A.26)}$$

Clearly, (A.26) is in the form of (A.1) whose solution is given by (A.5) in the form

$$G^T_G(x) = \sum_{r=0}^{[x/t_1]} G^T_r(x) \qquad \text{(A.27)}$$

where $\hat{G}^T_r(x)$ satisfies

$$\tilde{G}^T_r(x) = \frac{(-\mu(x - rt_1))^r}{r!} - \mu_2 \int_{\max(T,rt_1)}^{x-t_2} \hat{G}^T_r(y)\, dy \quad , \quad x \geq rt_1. \qquad \text{(A.28)}$$

Consider two cases, $T \geq rt_1$ and $rt_1 < T$. For the first case of $T \geq rt_1$, let $d^1_v(x)$ be as in (A.9) and (A.10) with

$$z^1(x) = \frac{(-\mu_1(x - rt_1))^r}{r!} \quad . \qquad \text{(A.29)}$$

Then

$$d^1_0(x) = \begin{cases} \dfrac{(-\mu_1(x - rt_1))^r}{r!} & , \text{ if } \quad x \geq rt_1 \\ 0 & , \text{ if } \quad x < rt_1 \end{cases} \qquad \text{(A.30)}$$

and by repeated use of (A.10)

$$d_v^1(x) = \begin{cases} \dfrac{(-\mu_1)^r\ (-\mu_2)^v\ (x - rt_1 - vt_2)^{r+v}}{(r+v)!}, & \text{if} \quad x \geq rt_1 + vt_2 \\ 0, & \text{if} \quad x < rt_1 + vt_2 \end{cases} \quad (A.31)$$

Applying Lemma A.1

$$\tilde{G}_r^T(x) = \sum_{v=0}^{[(x-T)/t_2]} d_v^1(x)$$

$$= \sum_{v=0}^{[(x-T)/t_2]} \frac{(-\mu_1)^r\ (-\mu_2)^v\ (x - rt_1 - vt_2)^{r+v}}{(r+v)!} \qquad (A.32)$$

for $T \geq rt_1$.

For the second case of $rt_1 < T$, let $d_v^2(x)$ be as in (A.9) and (A.10) with

$$z^2(x) = \frac{(-\mu_1(x - rt_1))^r}{r!} \quad . \qquad (A.33)$$

Using the binomial expansion, $z^2(x)$ is given by

$$z^2(x) = (\mu_1)^r \sum_{u=0}^{r} \left(\frac{(x - T)^u}{u!}\right) \frac{(T - rt_1)}{(r-u)!} \quad . \qquad (A.34)$$

So,

$$d_0(x) = \begin{cases} (-\mu_1)^r \displaystyle\sum_{u=0}^{r} \frac{(x - T)^u}{u!} \frac{(T - rt_1)}{(r-u)!}, & \text{if } x \geq T \\ 0, & \text{if } x < T \end{cases} \qquad (A.35)$$

and by repeated use of (A.10)

$$d_v^2(x) = (-\mu_1)^r\ (-\mu_2)^v \sum_{u=0}^{r} \left(\frac{(x - T - v\,t_2)^{v+u}}{(v+u)!}\right) \frac{(T - rt_1)^{r-u}}{(r-u)!} \quad . \qquad (A.36)$$

Applying Lemma A.1 to (A.28),

$$\tilde{G}_r^T(x) = \sum_{v=0}^{[(x-T)/t_2]} d_v^2(x)$$

$$= \sum_{v=0}^{[(x-T)/t_2]} (-\mu_1)^r (-\mu_2)^v \sum_{u=0}^{r} \frac{(x - T - vt_2)^{u+v}}{(u+v)!} \frac{(T - ut_1)^{r-u}}{(r-u)!} \tag{A.37}$$

for $rt_1 < T$.

Q.E.D

**Lemma A.4**

Let $G_e(x)$ satisfy

$$G_e(x) = e^{-\lambda x} - \mu_2 \int_0^{x-t_2} G_e(y)\, dy \quad , \quad x \geq 0 \tag{A.38}$$

where $G_e(x) = 0$ for $x < 0$, and $\lambda$, $\mu_2$ and $t_2$ are positive numbers.

Then

$$G_e(x) = \begin{cases} \sum_{r=0}^{[x/t_2]} (-\mu_2)^r \{(-\lambda)^{-r} e^{-\lambda(x-vt_2)} - \sum_{v=1}^{r} \frac{(x-rt_2)^{r-v}}{(r-v)!} (-\lambda)^{-v}\} & , \text{ if } \quad x \geq 0 \\ 0 & , \text{ if } \quad x < 0. \end{cases} \tag{A.39}$$

**Proof**

Let $d_r(x)$ be given by (A.9) and (A.10) where

$$z(x) = e^{-\lambda x} \quad . \tag{A.40}$$

Then

$$d_0(x) = \begin{cases} e^{-\lambda x} & , \text{ if } \quad x \geq 0 \\ 0 & , \text{ if } \quad x < 0 \end{cases} \tag{A.41}$$

and by repeated use of (A.10)

$$d_r(x) = \begin{cases} (-\mu_2)^r \; \{(-\lambda)^{-r} \; e^{-\lambda(x-rt_2)} - \sum\limits_{v=1}^{n} \dfrac{(x-rt_2)^{r-v}}{(r-v)!} (-\lambda)^{-v}\}, \text{ if } x \geq rt_2 \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad , \text{ if } x < rt_2. \end{cases} \tag{A.42}$$

Applying Lemma A.1, $G_e(x)$ is given by

$$G_e(x) = \sum_{r=0}^{[x/t_2]} d_r(x) \quad . \tag{A.43}$$

Hence, upon substitution of (A.42) into (A.43), (A.39) results.

<u>Q.E.D</u>

To insure the proper normalization of the solution of (4.24), various limits of the functions $G^T_{t_1}(x)$, $G^T_G(x)$, and $G_e(x)$ defined in Lemmas A.2, A.3, and A.4 need to be determined. The following lemmas determine the limits by using the final value theorem of Laplace transform theory.

<u>Lemma A.5</u>

Let $G^T_{t_2}(x)$ be given by (A.16) where

$$\mu_2 = \lambda e^{-\lambda t_2} , \tag{A.44}$$

and $\lambda t_2 < 1$. Then

$$\lim_{x\to\infty} e^{\lambda x} G_{t_2}^{T}(x) = \frac{e^{\lambda T}}{1 - \lambda t_2} . \tag{A.45}$$

<u>Proof</u>

Let $\hat{G}_{t_2}^{T}(s)$ be the Laplace transform of $e^{\lambda x} G_{t_2}^{T}(x)$. So, using (A.16)

$$\hat{G}_{t_2}^{T}(s) = e^{\lambda T} L\{ \sum_{r=0}^{[x-T/t_2]} \frac{(-\lambda(x-rt_2-T))^r}{r!} e^{\lambda(x-ra_2-T)} u(x-rt_2-T)\}$$

$$= e^{\lambda T} \sum_{r=0}^{\infty} L\{(\frac{-\lambda(x-rt_2-T)}{r!})^r e^{\lambda(x-ra_2-T)} u(x-rt_2-T)\} \tag{A.46}$$

where $u(x)$, (4.83), is the unit step function at zero. Using the relationship

$$L\{\frac{(x - A)^r}{r!} e^{\lambda(x-A)} u(x - A)\} = \frac{e^{-sA}}{(s - \lambda)^{r+1}} \tag{A.47}$$

in (A.46), results in

$$\hat{G}_{t_2}^{T}(s) = \sum_{r=0}^{\infty} \frac{e^{-T(s-\lambda}}{(s - \lambda)} \left(\frac{-\lambda e^{-st_2}}{s - \lambda}\right)^r .$$

$$= \frac{e^{-T(s-\lambda)}}{s - \lambda + \lambda e^{-st_2}} \tag{A.48}$$

Now, from the final value theorem of Laplace transforms,

$$\lim_{x\to\infty} e^{\lambda x} G_{t_2}^T(x) = \lim_{s\to\infty} s\,\hat{G}_{t_2}^T(s) \quad . \tag{A.49}$$

So, using (A.48) and l'Hospital rule in (A.49),

$$\lim_{x\to\infty} e^{\lambda x} G_{t_2}^T(x) = \frac{e^{\lambda T}}{1 - \lambda t_2} \quad . \tag{A.50}$$

Lemma A.6

Let $G_G^T(x)$ be given by (A.25) where

$$\mu_i = \lambda e^{-\lambda t_i} \tag{A.51}$$

and $\lambda t_2 < 1$. Then

$$\lim_{x\to\infty} e^{\lambda x}\{G_G^{T_1}(x) - G_G^{T_2}(x)\} = \sum_{r=0}^{[T_2/t_1]} (L_r^{T_1} - L_r^{T_2}) \tag{A.52}$$

where

$$L_r^T = \begin{cases} 1/(1-\lambda t_2) & , \quad \text{if } T \le rt_1 \\ \displaystyle\sum_{u=0}^{r} \frac{(-\lambda(T-rt_1))^u}{(1-\lambda t_2)} e^{\lambda(T-rt_1)} & , \quad \text{if } rt_1 < T \end{cases} \tag{A.53}$$

and $T_1 < T_2$.

Proof

Let $G_r^{T_1T_2}(x)$ be given by

$$G_r^{T_1T_2}(x) = \tilde{G}_r^{T_1}(x) - \tilde{G}_r^{T_2}(x) \tag{A.54}$$

where $\tilde{G}_r^T(x)$ is given by (A.25). Furthermore, let $\hat{G}_r^{T_1T_2}(s)$ be the Laplace transform of $e^{\lambda x}\, G_r^{T_1T_2}(x)$. Then, from (A.24), the Laplace transform of $e^{\lambda x}\, \tilde{G}_G^{T_1T_2}(x)$ is given by

$$L\{e^{\lambda x}(G_G^{T_1}(x) - G_G^{T_2}(x))\} = \sum_{r=0}^{\infty} \hat{G}_r^{T_1T_2}(s) \quad . \tag{A.55}$$

Taking the Laplace transform of $e^{\lambda x}\, G_r^{T_1T_2}(x)$ and interchanging the order of summation and transformation in the resulting expression and using (A.47), $\hat{G}_r^{T_1T_2}(s)$ is given by

$$\hat{G}_r^{T_1T_2}(s) = \begin{cases} 0 & , \text{ if } T_2 \le rt_1 \\ \displaystyle\sum_{v=0}^{\infty} (-\lambda)^{r+v} \left\{ \frac{e^{-s(rt_1+vt_2)}}{(s-\lambda)^{v+1}} - e^{\lambda(T_2-rt_1)} \right. & \\ \displaystyle\quad \left. \cdot \sum_{u=0}^{r} \frac{e^{-s(T_2+vt_2)}}{(s-\lambda)^{u+v+1}} \frac{(T_2 - rt_1)^{r-u}}{(r-u)!} \right\} & , \text{ if } T_1 \le rt_1 < T_2 \\ \displaystyle\sum_{v=0}^{\infty} (-\lambda)^{r+v} \sum_{u=0}^{r} \frac{e^{-(\lambda rt_1 - svt_2)}}{(s-\lambda)^{v+u+1}} \left\{ \frac{(T_1 - rt_1)^{r-u}}{(r-u)!} e^{-T_1(s-\lambda)} \right. & \\ \displaystyle\quad \left. - \frac{(T_2 - rt_1)}{(r-u)!} e^{-T_2(s-\lambda)} \right\} & , \text{ if } rt_1 < T_1 \, . \end{cases} \tag{A.56}$$

Exchanging the order of summation in (A.56),

$$\hat{G}_r^{T_1T_2}(s) = \begin{cases} 0 \quad , & \text{if } T_2 \le rt_1 \\ \dfrac{(-\lambda)^r}{s-\lambda+\lambda e^{-st_2}} \left\{ \dfrac{e^{-srt_1}}{(s-\lambda)^r} - \sum_{u=0}^{r} \dfrac{e^{-T_2(s-\lambda)}}{(s-\lambda)^u} \right. \\ \quad \left. \cdot \dfrac{(T_2 - rt_1)^{r-u}}{(r-u)!} e^{-\lambda rt_1} \right\} \quad , & \text{if } T_1 \le rt_1 < T_2 \\ \dfrac{(-\lambda)^r e^{-\lambda rt_1}}{s-\lambda+\lambda e^{-st_2}} \sum_{u=0}^{r} \dfrac{1}{(s-\lambda)^u} \left\{ \dfrac{(T_1 - rt_1)^{r-u}}{(r-u)!} e^{-T_2(s-\lambda)} \right. \\ \quad \left. - \dfrac{(T_2 - rt_1)^{r-u}}{(r-u)!} e^{-T_2(s-\lambda)} \right\} \quad , & \text{if } rt_1 < T_1 \end{cases} \quad . \quad (A.57)$$

Now, from the final value theorem of Laplace transforms and (A.55),

$$\lim_{x\to\infty} e^{\lambda x}(G_G^{T_1}(x) - G_G^{T_2}(x)) = \lim_{s\to 0} s\, L\{e^{\lambda x}(G_G^{T_1}(x) - G_G^{T_2}(x))\}$$

$$= \lim_{s\to 0} \sum_{r=0}^{\infty} s\, \hat{G}_r^{T_1T_2}(s) \quad . \quad (A.58)$$

But from (A.52) and l'Hospital rule,

$$\lim_{s\to 0} \hat{G}_r^{T_1T_2}(s) = \frac{1}{1 - \lambda t_2}(L_r^{T_1} - L_r^{T_2}) \quad (A.59)$$

where $L_r^T$ is given by (A.53). So, substituting (A.59) into (A.58), it is clear that

$$\lim_{x\to\infty} e^{\lambda x}(G_G^{T_1}(x) - G_G^{T_2}(x)) = \sum_{r=0}^{[T_2/t_1]} (L_r^{T_1} - L_r^{T_2}) \quad . \quad (A.60)$$

<u>Q.E.D</u>

**Lemma A.7**

Let $G_e(x)$ be given by (A.39), where

$$\mu_2 = \lambda e^{-\lambda t_2} \tag{A.61}$$

and $\lambda t_2 < 1$. Then

$$\lim_{x\to\infty} e^{\lambda x}(G_e(x) - e^{-\lambda T} G_e(x - T)) = -\frac{\lambda T}{1 - \lambda t_2} \tag{A.62}$$

where $T > 0$.

**Proof**

From (A.39)

$$e^{\lambda x}(G_e(x) - e^{-\lambda T} G_e(x - T))$$

$$= \sum_{r=0}^{[x/t_2]} \left\{1 - \sum_{v=1}^{r} \frac{(-\lambda(x - rt_2))^{r-v}}{(r-v)!} e^{\lambda(x-rt_2)}\right\}$$

$$- \sum_{r=0}^{[x-T/t_2]} \left\{1 - \sum_{v=1}^{r} \frac{(-\lambda(x - rt_2 - T))^{r-v}}{(r-v)!} e^{\lambda(x-rt_2-T)}\right\} . \tag{A.63}$$

Let $\hat{G}_e^T(s)$ be the Laplace transform of (A.63). Then by interchanging the order of summation and transformation in the expression for $\hat{G}_e^T(s)$ and using relationship (A.47), $\hat{G}_e^T(s)$ is given by

$$\hat{G}_e^T(s) = \sum_{r=0}^{\infty} e^{-srt_2}\left\{\frac{1}{s} - \frac{1}{s-\lambda} \sum_{v=0}^{n} \left(\frac{-\lambda}{s-\lambda}\right)^{r-v}\right\}(1 - e^{-sT})$$

$$= \frac{(s-\lambda)(1 - e^{-sT})}{s\left(s - \lambda + \lambda e^{-st_2}\right)} . \tag{A.64}$$

Now, from the final value of theorem of Laplace transforms and l'Hospital rule,

$$\lim_{x\to\infty} e^{\lambda x}(G_e(x) - e^{-\lambda T} G_e(x - T)) = \lim_{s\to 0} s\, \hat{G}_e^T(s)$$

$$= \frac{-\lambda T}{1 - \lambda t_2} \quad . \tag{A.65}$$

Q.E.D

A.3 Terms in the Evaluation of the Mean Waiting Time

In the evaluation of the mean waiting time in Theorem 4.3, it is required to evaluate the limits of specific functions. The following lemmas provide the limits to these functions.

**Lemma A.8**

Let $G_{t_2}^T(x)$ be given by (A.16) and let

$$H_{t_2}^T(x) = \int_0^x y\, d(e^{\lambda y} G_{t_2}^T(y)) \tag{A.66}$$

with $\lambda t_2 < 1$. Then

$$\lim_{x\to\infty} H_{t_2}^T(x) = e^{\lambda T}\left(\frac{T}{1 - \lambda t_2} + \frac{\lambda t_2{}^2}{2(1 - \lambda t_2)^2}\right) \quad . \tag{A.67}$$

**Proof**

The Laplace transform of $e^{\lambda x} G_{t_2}^T(x)$ which is denoted $\hat{G}_{t_2}^T(s)$ is given by (A.48). Clearly from (A.66) the transform of $H_{t_2}^T(x)$ is given

by

$$L(H_{t_2}^{T}(x)) = -\frac{1}{s}\frac{d}{ds}(s\,\hat{G}_{t_2}^{T}(s)) \quad . \qquad (A.68)$$

Using the final theorem of Laplace transforms and (A.68),

$$\lim_{x\to\infty} H_{t_2}^{T}(x) = \lim_{s\to 0} s\, L(H_{t_2}^{T}(x))$$

$$= -\lim_{s\to 0} \frac{d}{ds}[s\,\hat{G}_{t_2}^{T}(s)] \quad . \qquad (A.69)$$

Finally, substituting (A.48) in (A.69) and evaluating the limit in (A.69), the result (A.67) is proved.

Q.E.D

Lemma A.9

Let $G_G^T(x)$ be given by (A.24) and let

$$H_G^{T_1T_2}(x) = \int_0^x y\, d(e^{\lambda y}\{G_G^{T_1}(y) - G_G^{T_2}(y)\}) \qquad (A.70)$$

with $\lambda t_2 < 1$. Then for $T_1 < T_2$

$$\lim_{x\to\infty} H_G^{T_1T_2}(x) = \frac{1}{1-\lambda t_2}\sum_{r=0}^{[T_2/t_1]} (h_r^{T_1} - h_r^{T_2}) \qquad (A.71)$$

where

$$h_r^T = \begin{cases} \dfrac{2r(1 - \lambda t_2)(1 - \lambda t_1) + (\lambda t_2)^2}{2\lambda(1 - \lambda t_2)^2}, & \text{if } T \leq rt_1 \\ \displaystyle\sum_{u=0}^{r} \left(\frac{(-\lambda(T-rt_1))^u}{u!}\right) \frac{[2(1-\lambda t_2)(T\lambda+u-r) + (\lambda t_2)^2]\, e^{\lambda(T-rt_1)}}{2\lambda(1 - \lambda t_2)}, & \text{if } rt_1 < T \end{cases} . \tag{A.72}$$

Proof

From (A.55), the Laplace transform of $e^{\lambda x}(G_G^{T_1}(x) - G_G^{T_2}(x))$ is denoted $\sum_{r=0}^{\infty} \hat{G}_r^{T_1T_2}(s)$. In a similar manner to (A.68) and (A.69)

$$L(H_G^{T_1T_2}(x)) = -\frac{1}{s}\frac{d}{ds}\left(s \sum_{r=0}^{\infty} \hat{G}_r^{T_1T_2}(s)\right) \tag{A.73}$$

and

$$\lim_{x\to\infty} H_G^{T_1T_2}(x) = -\lim_{s\to 0}\frac{d}{ds}\left[s\left(\sum_{r=0}^{\infty} \hat{G}_r^{T_1T_2}(s)\right)\right] . \tag{A.74}$$

Then, subsituting (A.57) for $\hat{G}_r^{T_1T_2}(s)$ in (A.74) and evaluating the limit, (A.71) is obtained.

Q.E.D

<u>Lemma A.10</u>

Let $G_e(x)$ be given by (A.39) and let

$$H_e^{T_1T_2}(x) = \int_0^x y\, d\left(e^{\lambda x}\{G_e(x - T_1) - e^{-\lambda(T_2-T_1)} G_e(x - T_2)\}\right) \quad \text{(A.75)}$$

with $\lambda t_2 < 1$ and $T_1 < T_2$. Then

$$\lim_{x\to\infty} H_e^{T_1T_2}(x) = \frac{-\lambda(T_2^{\,2} - T_1^{\,2})}{2(1 - \lambda t_2)} - \frac{(T_2 - T_1)}{(1 - \lambda t_2)} - \frac{-(\lambda t_2)^2(T_2 - T_1)}{2(1 - \lambda t_2)^2} \quad . \quad \text{(A.76)}$$

<u>Proof</u>

From (A.64), the Laplace transform of $e^{\lambda x}(G_e(x) - e^{-\lambda T} G_e(x-T))$ is denoted $\hat{G}_e(s)$. In a similar manner to (A.68) and (A.69)

$$L(H_e^{T_1T_2}(x)) = -\frac{1}{s}\frac{d}{ds}\{s\, e^{-sT_1}\, G_e^{T_2-T_1}(s)\} \quad \text{(A.77)}$$

and

$$\lim_{x\to\infty} H_e^{T_1T_2}(x) = -\lim_{s\to 0}\frac{d}{ds}\{s\, e^{-sT_1}\, G_e^{T_2-T_1}(s)\} \quad . \quad \text{(A.78)}$$

Then, substituting (A.64) for $\hat{G}_e^T(s)$ in (A.78) and evaluating the limit, (A.76) is obtained.

<u>Q.E.D</u>